

Copyright (c) 1993 - 2000 Compugen Ltd.

Gencore version 4.5

**OM protein - protein search, using sw model**

Run on: July 12, 2001, 06:16:44 ; Search time 12.13 Seconds  
                   (without alignments)  
                   111.269 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 67

Sequence: 1 GGFGLGGKCPSEIFSR.....CRIGYLRNKKVCPRSKCG 67

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 193259 seqs, 2014635 residues

Word size : 0

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 50 summaries

Database : Issued\_Patents\_AA:\*

1: /cn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*

2: /cn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*

3: /cg2\_6/ptodata/2/1aa/6A\_COMB.pep:\*

4: /cn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*

5: /cn2\_6/ptodata/2/1aa/PCUTS\_COMB.pep:\*

6: /cn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Match Length	DB ID	Description
1	8	11.9	738	3 US-09-864-038A-3 Sequence 3, Appli
2	7	10.4	47	2 US-09-637-759B-400 Sequence 400, Appli
3	7	10.4	47	3 US-09-871-355B-400 Sequence 400, Appli
4	6	9.0	130	2 US-09-888-497-43 Sequence 43, Appli
5	6	9.0	130	5 PCT-US94-07926-43 Sequence 43, Appli
6	6	9.0	158	2 US-09-888-497-22 Sequence 22, Appli
7	6	9.0	158	5 PCT-US94-07926-22 Sequence 22, Appli
8	6	9.0	253	1 US-09-399-696-4 Sequence 4, Appli
9	6	9.0	253	1 US-09-399-696-118 Sequence 118, Appli
10	6	9.0	400	1 US-07-730-953-2 Sequence 10, Appli
11	6	9.0	488	2 US-09-926-692-10 Sequence 2, Appli
12	6	9.0	509	1 US-09-030-096-2 Sequence 2, Appli
13	6	9.0	568	4 US-09-160-655-2 Sequence 89, Appli
14	6	9.0	590	4 US-09-413-811-89 Sequence 76, Appli
15	6	9.0	591	4 US-09-413-814-76 Sequence 16, Appli
16	6	9.0	2182	2 US-08-478-26B-16 Sequence 5, Appli
17	6	9.0	3075	2 US-08-460-309-5 Sequence 2, Appli
18	6	9.0	3075	2 US-08-125-775-5 Sequence 2, Appli
19	5	7.5	5	3 US-09-026-633-2 Sequence 12, Appli
20	5	7.5	7	4 US-08-929-329-12 Sequence 4, Appli
21	5	7.5	8	4 US-09-128-50-4 Sequence 11, Appli
22	5	7.5	9	1 US-08-423-069-11 Sequence 11, Appli
23	5	7.5	9	2 US-08-317-844B-11 Sequence 41, Appli
24	5	7.5	9	4 US-08-1663-168C-41 Sequence 42, Appli
25	5	7.5	9	4 US-08-963-168C-42 Sequence 43, Appli
26	5	7.5	9	4 US-09-353-976-3 Sequence 44, Appli
27	5	7.5	9	4 US-09-353-976-5 Sequence 45, Appli

**ALIGNMENTS**

RESULT 1  
       US-09-864-038A-3  
       ; Sequence 3, Application US/08864038A  
       ; Patent No. 600192  
       ; GENERAL INFORMATION:  
       ; APPLICANT: Kunio NAKASHIMA et al.  
       ; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE, cDNA, VECTOR  
       ; TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID  
       ; COMPUTER READABLE FORM:  
       ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
       ; COMPUTER: IBM Compatible  
       ; OPERATING SYSTEM: Microsoft Windows 95  
       ; SOFTWARE: Word Perfect 6.1  
       ; CURRENT APPLICATION DATA:  
       ; APPLICATION NUMBER: US/09/0864,038A  
       ; FILING DATE: May 28, 1997  
       ; PRIORITY APPLICATION DATA:  
       ; APPLICATION NUMBER: JP 8-184459  
       ; FILING DATE: 15-July-1996  
       ; ATTORNEY/AGENT INFORMATION:  
       ; NAME: C. Bruce Hamburg  
       ; REGISTRATION NUMBER: 22,389  
       ; REFERENCE/DOCKET NUMBER: F-5610  
       ; TELECOMMUNICATION INFORMATION:  
       ; TELEPHONE: (212)986-23440  
       ; TELEFAX: (212)953-7733  
       ; INFORMATION FOR SEQ ID NO: 3:  
       ; SEQUENCE CHARACTERISTICS:  
       ; LENGTH: 738  
       ; TYPE: amino acid  
       ; TOPOLOGY: linear  
       ; MOLECULE TYPE: protein  
       ; ORIGINAL SOURCE: protein  
       ; ORGANISM: Pinctada fucata

FEATURE: mantis epithelial cell  
 NAME/KEY: peptide  
 LOCATION: from 1 to 738  
 IDENTIFICATION METHOD: E (by experiment)  
 US-08-864-038A-3

Query Match Similarity 11.9%; Score 8; DB 3; Length 738;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFGGIGGG 8  
 Db 493 GGFGGIGGG 500

RESULT 2  
 US-08-637-759B-400  
 Sequence 400, Application US/08637759B  
 Patent No. 5876931

GENERAL INFORMATION:  
 APPLICANT: David William Holden  
 TITLE OF INVENTION: Identification of Genes  
 NUMBER OF SEQUENCES: 501  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Patrea L. Pabst  
 STREET: 2800 One Atlantic Center  
 STREET: 1201 West Peachtree Street  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: USA  
 ZIP: 30309-3450

COMPUTER READABLE FORM:  
 COMPUTER TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, version #1.30

CURRENT APPLICATION DATA:  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08-637,759B  
 FILING DATE: 03-MAY-1996  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/GB95/02875  
 FILING DATE: 11-DEC-1995  
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
 NAME: Pabst, Patrea L.  
 REGISTRATION NUMBER: 31,284  
 REFERENCE DOCKET NUMBER: RPMS 101  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (404) 873-8794  
 TELEX: (404) 873-8795

INFORMATION FOR SEQ ID NO: 400:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 47 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 US-08-871-355A-400

RESULT 3  
 US-08-871-355A-400  
 Sequence 400, Application US/08881355A  
 Patent No. 6015669

GENERAL INFORMATION:  
 APPLICANT: David William Holden  
 TITLE OF INVENTION: Identification of Genes  
 NUMBER OF SEQUENCES: 501  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Patrea L. Pabst  
 STREET: 2800 One Atlantic Center  
 STREET: 1201 West Peachtree Street  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: USA  
 ZIP: 30309-3450

COMPUTER READABLE FORM:  
 COMPUTER TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, version #1.30

CURRENT APPLICATION DATA:  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08-871,355A  
 FILING DATE: 09-JUN-1997  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/GB95/02875  
 FILING DATE: 11-DEC-1995  
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
 NAME: Pabst, Patrea L.  
 REGISTRATION NUMBER: 31,284  
 REFERENCE DOCKET NUMBER: RPMS 101 CON  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (404) 873-8794  
 TELEX: (404) 873-8795

INFORMATION FOR SEQ ID NO: 400:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 47 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 US-08-871-355A-400

RESULT 4  
 US-08-888-497-43  
 Sequence 43, Application US/0888497  
 Patent No. 5972677

GENERAL INFORMATION:  
 APPLICANT: Tischfield, Jay A.  
 APPLICANT: Seilhamer, Jeffrey J.

TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide Sequences and Low Molecular Weight Amino Acid Sequences Having Internal Ribosome Binding Sites

TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences Having Internal Ribosome Binding Sites

TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide Sequences Having Internal Ribosome Binding Sites

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
 ADDRESSEE: Russell PA  
 STREET: 200 East Broward Boulevard  
 CITY: Fort Lauderdale

Query Match Similarity 10.4%; Score 7; DB 2; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SRCGRC 25  
 Db 37 SRCGRC 43

RESULT 5  
 US-08-888-497-43  
 Sequence 43, Application US/0888497  
 Patent No. 5972677

GENERAL INFORMATION:  
 APPLICANT: Tischfield, Jay A.  
 APPLICANT: Seilhamer, Jeffrey J.

TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide Sequences and Low Molecular Weight Amino Acid Sequences Having Internal Ribosome Binding Sites

TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences Having Internal Ribosome Binding Sites

TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide Sequences Having Internal Ribosome Binding Sites

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
 ADDRESSEE: Russell PA  
 STREET: 200 East Broward Boulevard

QY 19 SRCGRC 25  
 Db 37 SRCGRC 43

RESULT 6  
 US-08-637-759B-400

\* STATE: FL  
 \* COUNTRY: USA  
 ZIP: 33301  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT-US94/07926  
 FILING DATE: 15-JUL-1994  
 CLASSIFICATION:  
 PRIORITY APPLICATION DATA:  
 NAME: Manso, Peter J.  
 APPLICATION NUMBER: US/08/651,405  
 FILING DATE: 26-JUL-1993  
 APPLICATION NUMBER: US/08/888,497  
 FILING DATE: 26-JUL-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Manso, Peter J.  
 REFERENCE/DOCKET NUMBER: IN21044-5  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 305-527-2498  
 TELEFAX: 305-764-4996  
 INFORMATION FOR SEQ ID NO: 43:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 130 amino acids  
 STRANDEDNESS: single  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US94-07926-43

---

RESULT 5  
 Query Match 9.0%; Score 6; DB 2; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 0; Mismatches 0; Indels 0; Gaps 0;

Qy	5 GLGGRG 10	Db	29 GLGGRG 34

RESULT 6  
 Query Match 9.0%; Score 6; DB 5; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	5 GLGGRG 10	Db	29 GLGGRG 34

; Sequence 22, Application US/08888497  
 ; Patent No. 592677  
 GENERAL INFORMATION:  
 APPLICANT: Tischfield, Jay A.  
 TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide Sequences and Low Molecular Weight Amino Acid Sequences  
 TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites  
 NUMBER OF SEQUENCES: 44  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster & Russell PA  
 STREET: 200 East Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 33301  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/888,497  
 FILING DATE: 26-JUL-1993  
 CLASSIFICATION:  
 PRIORITY APPLICATION DATA:  
 NAME: Manso, Peter J.  
 APPLICATION NUMBER: US/08/651,405  
 FILING DATE: 26-JUL-1993  
 APPLICATION NUMBER: IN21044-5  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 305-527-2498  
 TELEFAX: 305-764-4996  
 INFORMATION FOR SEQ ID NO: 22:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 158 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

MOLECULE TYPE: protein  
 US-08-888-497-22

---

GENERAL INFORMATION:

APPLICANT: Seilhamer, Jeffrey J.  
 TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide Sequences and Low Molecular Weight Amino Acid Sequences

NUMBER OF SEQUENCES: 126  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/399,696  
 FILING DATE: 02-MAR-1995  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/156,671  
 FILING DATE: 22-NOV-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Smith, William M.  
 REGISTRATION NUMBER: 30,223  
 REFERENCE/DOCKET NUMBER: 15522-000710  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 326-2400  
 TELEFAX: (415) 326-2422  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 253 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-399-696-4

RESULT 7  
 Query Match 9.0%; Score 6; DB 2; Length 158;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PCT-US94-07926-22  
 Sequence 22, Application PC/PCTUS9407926  
 GENERAL INFORMATION:  
 APPLICANT: Tischfield, Jay A.  
 TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide Title of Invention: Sequences Having Internal Ribosome Binding Sites  
 NUMBER OF SEQUENCES: 44  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster & Russell PA  
 STREET: 200 East Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 33301

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US94/07926  
 FILING DATE: 15-JUL-1994  
 CLASSIFICATION:  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/097,354  
 FILING DATE: 26-JUL-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Manso, Peter J.  
 REGISTRATION NUMBER: 32,264  
 REFERENCE/DOCKET NUMBER: IN21044-5  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 305-527-2498  
 TELEFAX: 305-764-4996  
 INFORMATION FOR SEQ ID NO: 22:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 158 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US94-07926-22

---

RESULT 9  
 Query Match 9.0%; Score 6; DB 1; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-08-399-696-118  
 Sequence 118, Application US/08399696  
 PATENT NO. 5756669  
 GENERAL INFORMATION:  
 APPLICANT:  
 TITLE OF INVENTION: P53-BINDING POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME  
 NUMBER OF SEQUENCES: 126  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/399,696  
 FILING DATE: 02-MAR-1995  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/156,671  
 FILING DATE: 22-NOV-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Smith, William M.  
 REGISTRATION NUMBER: 30,223  
 REFERENCE/DOCKET NUMBER: 15522-000710  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 326-2400  
 TELEFAX: (415) 326-2422  
 INFORMATION FOR SEQ ID NO: 118:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 253 amino acids

RESULT 8  
 US-08-399-696-4  
 Sequence 4, Application US/08399696  
 PATENT NO. 5756669

Query Match Similarity 9.0%; Score 6; DB 1; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGIGGR 9  
 Db 199 GGIGGR 204

RESULT 10  
 US-07-730-953-2  
 Sequence 2, Application US/07730953  
 Patent No. 5288614

GENERAL INFORMATION:  
 APPLICANT: BODENMILLER, Heinz  
 TITLE OF INVENTION: METHOD FOR THE DETECTION OF MALIGNANT  
 TITLE OF INVENTION: METHOD FOR THE DETECTION OF MALIGNANT  
 NUMBER OF SEQUENCES: 2  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Nikaido, Marmelstein, Murray & Oram  
 STREET: 655 Fifteenth Street N.W. Suite 330  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005-5701

COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY DISK  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patientin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICANT NUMBER: US/07/730,953  
 FILING DATE: 19910723  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: DE P 39 42 999-7  
 FILING DATE: 21-DEC-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KITTS, Monica C.  
 REGISTRATION NUMBER: 36 105  
 REFERENCE/DOCKET NUMBER: P564-1119

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)638-5000  
 TELEFAX: (202)638-4810  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 400 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein

US-07-730-953-2

Query Match Similarity 9.0%; Score 6; DB 2; Length 488;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GGGRGK 11  
 Db 441 GGGRGK 446

RESULT 12  
 US-08-030-096-2  
 Sequence 2, Application US/08030096  
 Patent No. 5426041

GENERAL INFORMATION:  
 APPLICANT: Fabjanski, Steven F.  
 APPLICANT: Arnsen, Paul G.  
 TITLE OF INVENTION: BINARY CRYPTOCYTOTOXIC METHOD OF HYBRID  
 TITLE OF INVENTION: SEED PRODUCTION  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington, D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109

COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY DISK  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patientin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

RESULT 11  
 US-08-928-692-10  
 Sequence 10, Application US/08928692

APPLICATION NUMBER: US/08/030,096  
 FILING DATE: 22-MAR-1993  
 CLASSIFICATION: 800  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/556,917  
 FILING DATE: 20-JUL-1990  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: PCT/CA91/00255  
 REGISTRATION NUMBER: 29,768  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REFERENCE/DOCKET NUMBER: 33229/164/PITHI  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 509 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-030-096-2

Query Match 9.0%; Score 6; DB 1; Length 509;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGFGL 6  
 Db 94 GGFGL 99

RESULT 13  
 US-09-160-065-2  
 ; Sequence 2, Application US/09160065  
 ; Patent No. 6146641  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lee, Lucy  
 ; APPLICANT: Fadly, Aly  
 ; APPLICANT: Hunt, Henry  
 ; TITLE OF INVENTION: Avian Leukosis Virus Subgroup J Envelope Gene Product  
 ; TITLE OF INVENTION: For Diagnosis and Vaccine  
 ; FILE REFERENCE: Docket No. 6146641 0155.98 - Lee, Lucy  
 ; CURRENT APPLICATION NUMBER: US/09/160,065  
 ; CURRENT FILING DATE: 1998-09-24  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 568  
 ; TYPE: PRT  
 ; ORGANISM: Avian leukosis virus  
 US-09-160-065-2

Query Match 9.0%; Score 6; DB 1; Length 509;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 VCVPRS 64  
 Db 145 VCVPRS 150

RESULT 13  
 US-09-160-065-2  
 ; Sequence 2, Application US/09160065  
 ; Patent No. 6146641  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bristol-Myers Squibb, Co.  
 ; APPLICANT: Beyer, Stefan  
 ; APPLICANT: Beyer, Helmut  
 ; APPLICANT: Blöcker, Petra  
 ; APPLICANT: Ciro, Paul M  
 ; APPLICANT: Dougherty, Brian A  
 ; APPLICANT: Goldberg, Steven L  
 ; APPLICANT: Hofle, Gerhard  
 ; APPLICANT: Mueller, Joachim  
 ; APPLICANT: Reichenbach, Hans  
 ; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
 ; TITLE OF INVENTION: heteropolyketide compounds  
 ; FILE REFERENCE: PCT/US 99/23535  
 ; CURRENT APPLICATION NUMBER: DE 198 46 493.2  
 ; CURRENT FILING DATE: 1998-10-09  
 ; EARLIER FILING DATE: 1998-10-09  
 ; NUMBER OF SEQ ID NOS: 107  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 89  
 ; LENGTH: 590  
 ; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum  
 US-09-413-814-89

Query Match 9.0%; Score 6; DB 4; Length 590;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 VCVPRS 64  
 Db 145 VCVPRS 150

RESULT 15  
 US-09-413-814-76  
 ; Sequence 76, Application US/09413814  
 ; Patent No. 6225064  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
 ; APPLICANT: Bristol-Myers Squibb, Co.  
 ; APPLICANT: Beyer, Stefan  
 ; APPLICANT: Blöcker, Helmut  
 ; APPLICANT: Brandt, Petra  
 ; APPLICANT: Ciro, Paul M  
 ; APPLICANT: Dougherty, Brian A  
 ; APPLICANT: Goldberg, Steven L  
 ; APPLICANT: Hofle, Gerhard  
 ; APPLICANT: Mueller, Joachim  
 ; APPLICANT: Reichenbach, Hans  
 ; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
 ; TITLE OF INVENTION: heteropolyketide compounds  
 ; FILE REFERENCE: PCT/US 99/23535  
 ; CURRENT APPLICATION NUMBER: US/09/0413,814  
 ; CURRENT FILING DATE: 1999-10-07  
 ; EARLIER FILING DATE: 1998-10-09  
 ; NUMBER OF SEQ ID NOS: 107  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 76  
 ; LENGTH: 591  
 ; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum  
 US-09-413-814-76

Query Match 9.0%; Score 6; DB 4; Length 591;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 VCVPRS 64  
 Db 145 VCVPRS 151

RESULT 14  
 US-09-413-814-89  
 ; Sequence 89, Application US/09413814  
 ; Patent No. 6225064  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
 ; APPLICANT: Bristol-Myers Squibb, Co.



APPLICATION NUMBER: US/08/125,077  
FILING DATE: 22-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US 94/10730  
FILING DATE: 21-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/472,319  
FILING DATE: 30-JAN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/919,951  
FILING DATE: 27-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LA 9721  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3075 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

US-08-125-077-5

APPLICANT: Dalton, John P  
ATTORNEY: Kappe, Stefan  
TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing Vaccine Compositions  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Barnes & Thornburg  
STREET: 11 S Meridian  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

```

Query Match 9.0%; Score 6; DB 2; length 3075;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0;
Gaps 0
Oy 46 CVCRLG 51
Db 1528 CVCRLG 1533

```

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: Peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal

PATENT NO. 6053323  
GENERAL INFORMATION:  
APPLICANT: McMorris, Trevor C.  
APPLICANT: Keline, Michael J.  
TITLE OF INVENTION: Antitumor agent  
FILE REFERENCE: 103-008US1  
CURRENT APPLICATION NUMBER: US-A9/  
CURRENT FILING DATE: 1998-02-20  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: fastSEQ for Windows Vers  
SEQ ID NO 2  
LENGTH: 5  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Amino acid seq  
US-09-026-633-2

Query	Match	Score	Length	DB	Best Local Similarity	Pred.	No.	Mismatches	Indels	Gaps
Qy	5 GLGGR	9	7	7	100.0%	0	1.4e+05	0	0	0
Db	1 GLGGR	5								

RESULT 21  
US-09-128450-4  
Sequence 4, Application US/09128450  
; Patent No. 621119  
; GENERAL INFORMATION:  
APPLICANT: Chesbro, Bruce W  
INVENTOR: Chesbro, Bruce W

Query Match		7.5%	Score
Best	Local	Similarity	100.0%
Matches		Conservative	0
Qy	21	CDGR <sub>C</sub>	25
Db	1	CDGR <sub>C</sub>	5
RESULT	20		
US-08-979-329-12			
; Sequence 12, Application			US/08929329
; Patent No. 612070			
; GENERAL INFORMATION:			
; APPLICANT: Adams, John H			



CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/963,168C  
FILING DATE: 03-NOV-1997

ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, Peter J.  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 07117/059001

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 41:

SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

FEATURE:  
LOCATION: 3..3  
OTHER INFORMATION: xaa = any amino acid

8-963-168C-41

Query Match 7.5%; Score 5; DB 4; Length 9;  
Best Local Similarity 100%; Pred. No. 1.4e+0;  
Matches 5; Conservative 0; Mismatches 0;  
Indels 0; Gaps 0;

LT 25  
8-963-168C-42  
SEQUENCE 42, Application US/08963168C  
tent No. 6127166

GENERAL INFORMATION:  
APPLICANT: Bayley, Hagan  
APPLICANT: Cao, Quiping  
APPLICANT: Wang, YunLauN  
TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES  
TITLE OF INVENTION: AND GENES ENCODING THEM  
NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/963,168C  
FILING DATE: 03-NOV-1997

ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, Peter J.  
REGISTRATION NUMBER: 32,983

REFERENCE/DOCKET NUMBER: 07117/059001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

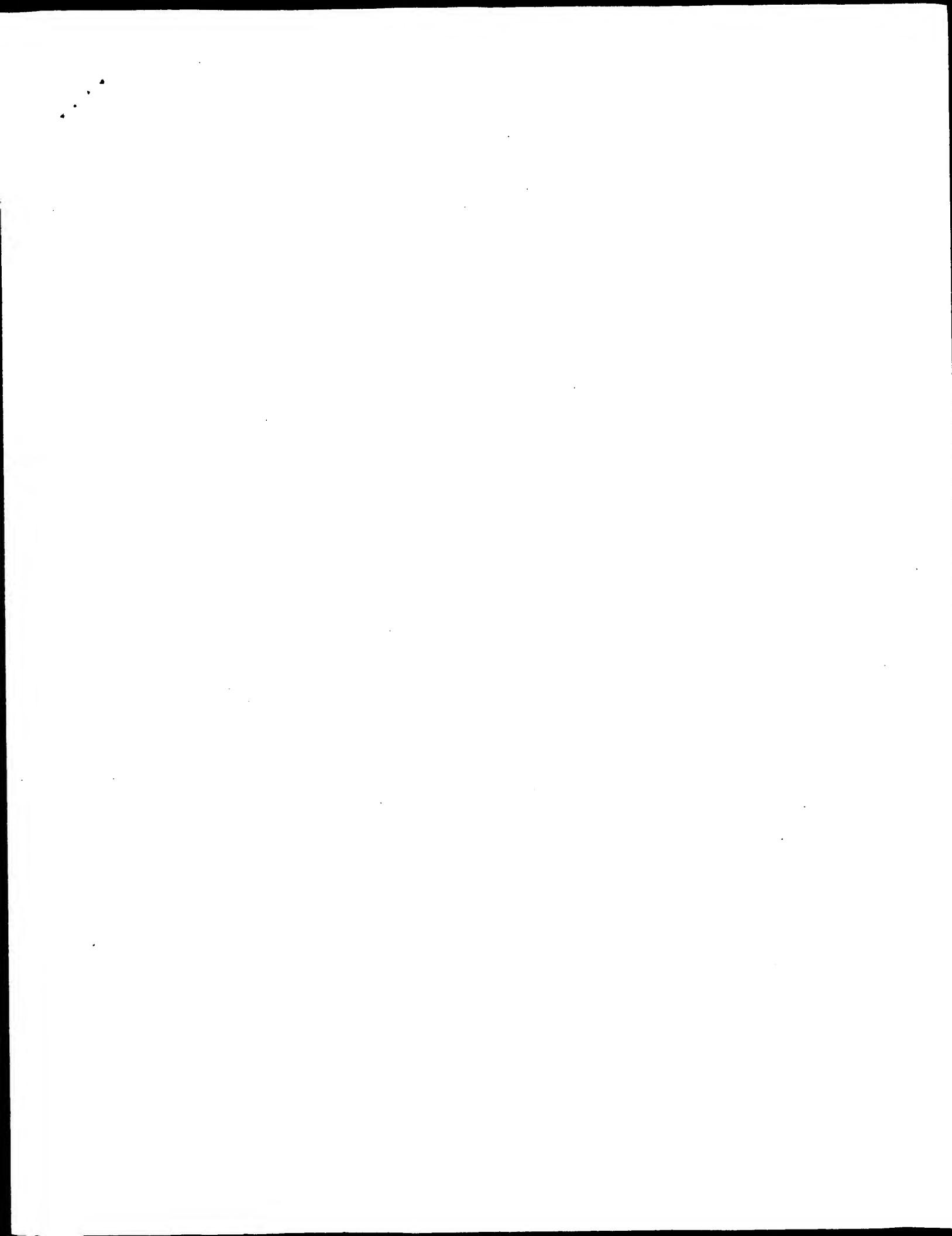
FEATURE:  
LOCATION: 3..3  
OTHER INFORMATION: xaa = any amino acid

US-08-963-168C-42

Search completed: July 12, 2001, 06:17:06  
Job time: 22 sec

Thu Jul 12 06:30:34 2001

us-09-506-978-1.oligo.rai



Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

OM protein - protein search, using sw model

Run on:

	July 12, 2001, 06:17:09 ; Search time 20.97 Seconds
	(without alignments)
	422.720 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 67

Sequence: 1 GFGESGLGGRKKCPSEIFR.....CRQSYLIRNKKKVCFVRSKCG 67

Scoring table: Oligo

Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 122305027 residues

Word size : 0

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database : SPTREMBL\_16;\*

- 1: sp\_archaea;\*
- 2: sp\_bacteria;\*
- 3: sp\_fungi;\*
- 4: sp\_human;\*
- 5: sp\_invertebrate;\*
- 6: sp\_mammal;\*
- 7: sp\_mhc;\*
- 8: sp\_organelle;\*
- 9: sp\_phage;\*
- 10: sp\_plant;\*
- 11: sp\_rodent;\*
- 12: sp\_unclassified;\*
- 13: sp\_vertebrate;\*
- 14: sp\_virus;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	8	11.9	113	5 Q9V5U5
2	8	11.9	151	10 Q92600
3	8	11.9	738	5 Q02402
4	8	11.9	904	5 Q76271
5	8	11.9	922	5 Q044367
6	8	11.9	1329	5 Q06810
7	7	10.4	100	10 Q95066
8	7	10.4	107	10 Q9LRP7
9	7	10.4	284	5 Q2103
10	7	10.4	313	5 Q9NL74
11	7	10.4	314	5 Q76140
12	7	10.4	438	5 Q9RY66
13	7	10.4	450	2 Q83238
14	7	10.4	463	2 Q10787
15	7	10.4	465	2 Q50943
16	7	10.4	556	2 Q9LRK4
17	7	10.4	620	10 Q9RFU14
18	7	10.4	652	10 Q68071
19	7	10.4	682	5 Q22537

RESULT 1

ID	Q9V5U5	PRELIMINARY:	PRT;	113 AA.
AC	Q9V5U5;			
DT	01-MAY-2000	(TREMBrel. 13, Created)		
DT	01-MAY-2000	(TREMBrel. 13, Last sequence update)		
DT	01-MAY-2000	(TREMBrel. 13, Last annotation update)		
DE	CG13217	PROTEIN.		
GN	CG13217.			
OS	Drosophila melanogaster	(Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Proteroggota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydriidae; Drosophilidae; Drosophila.			
OX	NCBI_TAXID=7271;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RX	MEDLINE-201506; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers J.Y.-H.C., Blazej R.G., Champe M., Pfleiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brockton P., Brostoller P., Burtsis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra T., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mayz A.D., Dew I., Dietz S.M., Dobson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin R.K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fosler C., Gabrilian A.E., Garg N.S., Gebhart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman-T.J., Hernandez J.R., Hoick J., Hoskin D., Houston K.A., Howland T.J., Wei M.-H., Ikegami C.,			

ALIGNMENTS



RC TISSUE\_FOOT;  
 RX MEDLINE=98070444; PubMed=9405478;  
 RA Qin X.X., Coyne K.J., Waite J.H.;  
 RP "Tough tendons. Mussel byssus has collagen with silk-like domains.";  
 RL J. Biol. Chem. 272:32623-32627 (1997).  
 DR EMBL: AF09249; AAB96638.1; .  
 DR InterPro: IPR000087; -.  
 DR Pfam: PF01391; Collagen; 7.  
 SQ SEQUENCE 922 AA; 80306 MW; 599D155P4TA2C24A CRC64;

Query Match 11.9%; Score 8; DB 5; Length 922;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFGGLGG 8  
 |||||||  
 Db 132 GGFGGLGG 139

RESULT 6  
 PRELIMINARY; PRT; 1329 AA.

ID 006810 DT 01-MAY-2000 (TREMBREL. 13, Created)  
 AC 006810 DT 01-MAY-2000 (TREMBREL. 13, Last sequence update)  
 DT 01-JUL-1997 (TREMBREL. 04, Created)  
 DT 01-MAR-2001 (TREMBREL. 16, Last annotation update)  
 DE HYPOTHETICAL 107.4 KDA PROTEIN.  
 GN PGRS\_FAMILY OR RV1450C OR MTG2493.04.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Actinomycetidae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TAXID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=9825987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglemeier K., Gas S., Barry C.E. III, Tekla F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Relman D.A., Rogers J.,  
 RA Relman D.A., Skarlev S., Squares R., Sulston J.E.,  
 RA Rutter S., Seeger K., Skelton S.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence. ";  
 RT Nature 393:537-544 (1998);  
 RL Nature 393:537-544 (1998);  
 DR EMBL: Z9544; CAB09271.1; .  
 DR Tuberculist; RV150c; .  
 DR InterPro: IPR00084; -.  
 DR InterPro: IPR002173; -.  
 DR InterPro: IPR003336; -.  
 DR Pfam: PF00934; PE: 1.  
 DR PR01370; TRANSITMINR.  
 DR PRODOM: PD001223; -.1.  
 DR PROSITE: PS00583; PFKB\_KINASES\_1; UNKNOWN\_2.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 1329 AA; 10740 MW; 777A125F6DBAB234 CRC64;

Query Match 11.9%; Score 8; DB 2; Length 1329;  
 Best Local Similarity 100.0%; Pred. No. 4.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFGGLGG 8  
 |||||||  
 Db 1318 GGFGGLGG 1325

RESULT 7  
 PRELIMINARY; PRT; 100 AA.

QSD66 ID QSD66

RESULT 9  
 PRELIMINARY; PRT; 284 AA.

AC 09SD66;  
 DT 01-MAY-2000 (TREMBREL. 13, Created)  
 DT 01-MAY-2000 (TREMBREL. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBREL. 13, Last annotation update)  
 DE HYPOTHETICAL 10.5 KDA PROTEIN.  
 GN F1112.120.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophytina;  
 OC Macrophyta; eudicotyledons; core eudicots; Rosidae; eurosids III;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Choise N., Robert C., Brottier P., Wincker P., Cattolico L.,  
 RA Arriguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X., Quetier F., Salanoubat M.;  
 RA Submitted (NOV-1999) to the EMBL/GenBank/DDJB databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RA Submitted (JAN-2000) to the EMBL/GenBank/DDJB databases.  
 DR EMBL: AL133292; CAB61953.1; .  
 SQ SEQUENCE 100 AA; 10530 MW; 9D9819F0BE5EDBD CRC64;

Query Match 10.4%; Score 7; DB 10; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFGGLGG 7  
 |||||||  
 Db 89 GGFGGLG 95

RESULT 8  
 PRELIMINARY; PRT; 107 AA.

Q9LEP7 ID Q9LEP7  
 PRELIMINARY; PRT; 107 AA.

AC 09LEP7; DT 01-OCT-2000 (TREMBREL. 15, Created)  
 DT 01-OCT-2000 (TREMBREL. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBREL. 16, Last annotation update)  
 DE PUTATIVE GLYCINE-RICH PROTEIN.  
 OS Brassica napus (Rape).  
 OC Eukaryota; Viridiplantae; Embryophyta; Spermatophytina;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Brassica.  
 OC NCBI\_TAXID=3707;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV\_N-O-9;  
 RA Bowers N.L., Trick M.;  
 RT "Microsynteny at the FCA region between Arabidopsis thaliana and  
 RT Brassica napus";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDJB databases.  
 DR EMBL: AJ23326; CAB01931.1; .  
 DR InterPro: IPR01525; -.1.  
 DR PROSITE; PS00095; C5\_MTASE\_2; UNKNOWN\_1.  
 DR SEQUENCE 107 AA; 9714 MW; 0A7BE4A01B09B1B3 CRC64;

Query Match 10.4%; Score 7; DB 10; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGFGGLGG 8  
 |||||||  
 Db 67 GGFGGLG 73

RESULT 9  
 PRELIMINARY; PRT; 284 AA.

Q21073 ID Q21073



RA Fraser C.M.;  
 "Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.;"  
 Science 286:1571-1577(1999).  
 EMBL: AE001871; AAF05676.1; -.  
 DR TIGR: DR084; -.  
 DR InterPro; IPR000765; -.  
 DR InterPro; IPR001684; -.  
 DR Pfam; PF01018; GTP1\_OBGB; 1.  
 DR PRINTS; PR00326; GTP1\_OBGB.  
 DR PRODOM; PD003114; -; 1.  
 DR PROSITE; PS00905; GTP1\_OBGB; 1.  
 SQ SEQUENCE 438 AA; 47836 MW; D0DC5612543F2989 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 20; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGLGGRG 10  
 |||||  
 Db 120 GGLGGRG 126

RESULT 13  
 083238 PRELIMINARY; PRT; 450 AA.  
 RC STRAIN=NICHOLS;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
 Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 McDonald L., Artiach P., Bowman C., Cottrell M.D., Fujii C., Garland S.,  
 Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 Venter J.C.,  
 RA "Complete genome sequence of *Treponema pallidum*, the syphilis  
 RT spirochete.;"  
 RL Science 281:375-389(1998).  
 DR EMBL: AE001203; AAC65198.1; -.  
 TIGR: TP0208; -.  
 DR InterPro; IPR001220; -.  
 DR InterPro; IPR001525; -.  
 DR InterPro; IPR002208; -.  
 DR Pfam; PF00344; secY; 1.  
 DR PRINTS; PR00303; SECYTRICASE.  
 DR PROSITE; PS00095; C5\_MTASe\_2; UNKNOWN\_1.  
 DR PROSITE; PS00307; LECTIN\_LEGUME\_BETA; UNKNOWN\_1.  
 DR PROSITE; PS00756; SECY\_2; UNKNOWN\_1.  
 SQ SEQUENCE 450 AA; 50265 MW; 892BF217FEB6E7E CRC64;

Query Match 10.4%; Score 7; DB 2; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 20; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRGK 11  
 |||||  
 Db 427 GLGGRGK 433

RESULT 14  
 083238 PRELIMINARY; PRT; 463 AA.  
 RC STRAIN=R37RV;  
 RX Murphy L., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -; SIMILARITY: TO UMUC, SAMB, AND MTCB PROTEINS.  
 DR EMBL: 274020; CAA98325.1; -.  
 DR Tuberculist; RV1537; -.  
 DR InterPro; IPR001126; -.  
 DR Pfam; PF00817; IMS; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 463 AA; 49107 MW; 43085A44CE8CC637 CRC64;  
 RP SEQUENCE FROM N.A.  
 RN [1]

RESULT 15  
 050943 PRELIMINARY; PRT; 465 AA.  
 ID 050943  
 AC 050943;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 Matches 7; Conservative 100.0%; Pred. No. 21; Mismatches 0; Indels 0; Gaps 0;  
 DT 01-NOV-1998 (TREMBLrel. 06, Last sequence update)  
 DE HYPOTHETICAL 52.8 KDA PROTEIN.  
 DR BRAS01  
 GN Borrelia burgdorferi (Lyme disease spirochete).  
 OS Plasmid lp54.  
 OC Bacteria; Spirochaetales; spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]

RESULT 15  
 050943 PRELIMINARY; PRT; 465 AA.  
 ID 050943  
 AC 050943;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 Matches 7; Conservative 100.0%; Pred. No. 21; Mismatches 0; Indels 0; Gaps 0;  
 DT 01-NOV-1998 (TREMBLrel. 06, Last sequence update)  
 DE HYPOTHETICAL 52.8 KDA PROTEIN.  
 DR BRAS01  
 GN Borrelia burgdorferi (Lyme disease spirochete).  
 OS Plasmid lp54.  
 OC Bacteria; Spirochaetales; spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;

Query Match 10.4%; Score 7; DB 2; Length 463;  
 Best Local Similarity 100.0%; Pred. No. 21; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRGK 11  
 |||||  
 DR EMBL: AE000790; AAC66282.1; -.  
 TIGR: BBA50; -.  
 KW Hypothetical protein; Plasmid.

SQ SEQUENCE 465 AA; 52773 MW; 8223EBE38B7E5B7 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 465;  
 Best Local Similarity 100.0%; Pred. No. 21; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRGK 11  
 |||||  
 DR EMBL: AE000790; AAC66282.1; -.  
 TIGR: BBA50; -.  
 KW Hypothetical protein; Plasmid.

QY 2 GRGGGG 8  
||| |||  
Db 44 GRGGGG 50

RESULT 16

Q918K4 PRELIMINARY; PRT; 556 AA.  
ID Q918K4;  
AC Q918K4;  
DT 01-OCT-2000 (TREMBREL\_15, Created)  
DT 01-OCT-2000 (TREMBREL\_15, Last sequence update)  
DT 01-MAR-2001 (TREMBREL\_16, Last annotation update)  
DE PE-PGRS HOMOLOG MAG24-2.  
GN MAG24-2.  
OS Mycobacterium marinum.  
OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TAXID=1781;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M  
RA Ramakrishnan L.; Federspiel N.A.; Falkow S.; Glycine-Rich PE-PGRS Family  
RT "A Mycopacterial Protein of the Repetitive Sequence and Contributors to Virulence."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL: AF201682; AAC65169.1; -  
DR InterPro: IPR00084; -  
DR InterPro: IPR001412; -  
DR Pfam: PF00934; PE; 1.  
DR PRODOM: PD001223; -; 1.  
DR PROSITE: PS00178; AA TRNA\_LIGASE\_I; UNKNOWN\_1;  
SQ SEQUENCE: 556 AA; 54576 MW; 8C812B5D368A634 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 556;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGUGGRG 10  
||| |||||  
Db 181 GGUGGRG 187

RESULT 17

Q9FU14 PRELIMINARY; PRT; 620 AA.  
ID Q9FU14;  
AC Q9FU14;  
DT 01-MAR-2001 (TREMBREL\_16, Created)  
DT 01-MAR-2001 (TREMBREL\_16, Last sequence update)  
DE PUTATIVE RECEPTOR KINASE.  
GN P0698G03\_13.

OS Oryza sativa (Rice).  
OC Buxaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;  
OC Oryza;  
OX NCBI\_TAXID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV\_NIPPONBARE;  
RA Sasaki T.; Matsumoto T.; Yamamoto K.;  
RT "Oryza sativa nippobare(GA3)" genomic DNA, chromosome 1, PAC  
clone P0698G03\_13;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL: AP002747; BAB17332.1; -  
DR Receptor; Kinase.  
SQ SEQUENCE: 620 AA; 69544 MW; C95560BF0BC1ABC6 CRC64;

Query Match 10.4%; Score 7; DB 10; Length 620;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GRGKCPs 14  
|||||||  
Db 30 GRGKCPs 36

RESULT 18

Q968071 PRELIMINARY; PRT; 652 AA.  
ID Q968071;  
AC 068071;  
DT 01-AUG-1998 (TREMBREL\_07, Created)  
DT 01-AUG-1998 (TREMBREL\_07, Last sequence update)  
DT 01-MAR-2001 (TREMBREL\_16, Last annotation update)  
DE DNA GYrase Subunit B (EC 5.98.1.3).  
DN Rhodobacter capsulatus (Rhodopseudomonas capsulata).  
OC Rhodobacter capsulatus (Rhodopseudomonas capsulata).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Rhodobacter.  
OX NCBI\_TAXID=1061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SB1003;  
RX MEDLINE=97404404; PubMed=9256491;  
RA Vicek C.; Paces V.; Malisev N.; Paces J.; Haselkorn R.; Ronstein M.;  
RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997); PASSAGE AND REJOINING OF DOUBLE-STRANDED DNA.  
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE AND REJOINING OF DOUBLE-STRANDED DNA.  
CC -1- SIMILARITY: TO DNA TOPOISOMERASE II FAMILY.  
DR EMBL: AF010496; AAC16157.1; -  
DR HSSP: P06982; LAJ6;  
DR InterPro: IPR000410; -  
DR InterPro: IPR001241; -  
DR InterPro: IPR002288; -  
DR InterPro: IPR002936; -  
DR Pfam: PF00986; DNA\_gyrase\_B\_C; 1.  
DR Pfam: PF01751; TopoIin\_1.  
DR PRINTS: PR00418; TP12\_FAMILY.  
DR PRODOM: PD000616; -; 1.  
DR PRODOM: PD149633; -; 1.  
DR PROSITE: PS00177; TOPOISOMERASE\_II; 1.  
DR SMART: SM00433; TOPAC; 1.  
DR ATP-binding\_Isomerase; Topoisomerase.  
SQ SEQUENCE: 652 AA; 71524 MW; 4CE26874F4289D CRC64;

Query Match 10.4%; Score 7; DB 2; Length 652;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GLGGRGK 11  
||| |||||  
Db 573 GLGGRGK 579

RESULT 19

Q22537 PRELIMINARY; PRT; 682 AA.  
ID Q22537;  
AC Q22537;  
DT 01-NOV-1996 (TREMBREL\_01, Created)  
DT 01-NOV-1996 (TREMBREL\_01, Last sequence update)  
DT 01-NOV-1998 (TREMBREL\_08, Last annotation update)  
DE SIMILARITY TO COLLAGEN.  
GN T17H7\_1.  
OS Caenorhabditis elegans.  
OC Buxaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peleoderrinae; Caenorhabditis.  
OX NCBI\_TAXID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL\_N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R.; Ainscough R.; Anderson K.; Baynes C.; Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hiller L., Jelier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 Parsons J., Percy C., Riffen L., Roopra A., Saunders D., Showroneen R.,  
 RA Smaldo N., Smith A., Sonhammar E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sprout J., Wolildman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans";  
 RL Nature 368:32-38(1994).  
 RL [1]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Favello A.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; U42841; AAC41701; .  
 SQ SEQUENCE 682 AA; 72150 MW; 73933507D23E98A2 CRC64;

Query Match 10.4%; Score 7; DB 5; Length 682;  
 Best Local Similarity 100.0%; Pred. No. 30; 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 GGLGGRG 10  
 Db 587 GGLGGRG 593

RESULT 20

Q9H6R3 PRELIMINARY; PRT; 686 AA.  
 AC Q9H6R3  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE CDNA: FL021963 FIS; CLONE HEP05583.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TAXID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikeda Y., Okamoto S.,  
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/Genbank/DDBJ databases.  
 DR EMBL; AK025616; BAB15190.1; -;  
 SQ SEQUENCE 686 AA; 74778 MW; 3386338C6FFFFD7E8 CRC64;

Query Match 10.4%; Score 7; DB 4; Length 686;  
 Best Local Similarity 100.0%; Pred. No. 30; 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 GGHGGRG 10  
 Db 44 GGHGGRG 50

RESULT 21

087472 PRELIMINARY; PRT; 753 AA.  
 AC 087472;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DE TRPL; 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

OS Treponema pallidum (subsp. pertenue) (Yaws treponeme).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 RN [1] NCBI\_TAXID=160;

Query Match 10.4%; Score 7; DB 2; Length 756;  
 Best Local Similarity 100.0%; Pred. No. 32; 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GFGGAGG 8  
 Db 169 GFGGGAGG 175

RESULT 23

007894 PRELIMINARY; PRT; 758 AA.  
 AC 007894;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-JUL-2000 (TREMBLrel. 13, Last annotation update)  
 DE MAJOR OUTER SHEATH PROTEIN HOMOLOG MSP.  
 GN MSP OR TRP0621;  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 RN [1] NCBI\_TAXID=160;

OX NCBI\_TAXID=168;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RA Stamm L.V., Greene S.R., Bergen H.L., Hardham J.M., Barnes N.Y.;  
 RA "Identification and Sequence Analysis of Treponema pallidum tprJ, a Member of a Polymorphic Multigene Family";  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 DR EMBL; AF073527; AAC83239.1; -;  
 SQ SEQUENCE 753 AA; 80692 MW; BB66C99841BDF6A8 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 753;  
 Best Local Similarity 100.0%; Pred. No. 32; 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GFGGAGG 8  
 Db 166 GFGGGAGG 172

RESULT 22

083337 PRELIMINARY; PRT; 756 AA.  
 AC 083337  
 DT 01-NOV-1998 (TREMBLrel. 03, Created)  
 DT 01-NOV-1998 (TREMBLrel. 03, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE PTP PROTEIN G (TPRG).  
 TP0317  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 RN NCBI\_TAXID=160;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NICHOLS;  
 RX MEDLINE=98322770; PUBMED=9665876;  
 RA Fraser C.M., Morris S.J., Weinstock G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwynn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Ariach P., Bowman C., Cotton M., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Ventter J.C.,  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete.";  
 RL Science 283:375-388(1998);  
 DR EMBL; AE001211; AAC65301.1; -;  
 DR TIGR; TP0317; -;  
 SQ SEQUENCE 756 AA; 81324 MW; 4D1FF6B80CE51738A CRC64;

Query Match 10.4%; Score 7; DB 2; Length 756;  
 Best Local Similarity 100.0%; Pred. No. 32; 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GFGGAGG 8  
 Db 169 GFGGGAGG 175

RP SEQUENCE FROM N.A.  
 RC STRAIN=NICHOOLS;  
 RA Hardham J.M.;  
 RL Thesis (1995), Unknown institution.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NICHOOLS;  
 RA Stamm L.V., Barnes N.Y., Frye J.F.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NICHOOLS;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwynn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 McDonald L., Artiach P., Bowman C., Cottont M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of *Treponema pallidum*, the syphilis  
 RT spirochete.";  
 RL Science 281:375-388(1998).  
 DR EMBL; 080957; AAC8171.1; -.  
 DR EMBL; AED01237; AAC65595.1; -.  
 DR TIGR; TP0621; -.  
 SQ SEQUENCE 758 AA; 81369 MW; FFF93A8951E261B8 CRC64;

Query Match 10.4%; Score 7; DB 2; Length 758;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGG3GGG 8  
 |||||  
 Db 169 GFGGGGG 175

RESULT 24  
 Q9KHF2 PRELIMINARY; PRT; 758 AA.  
 ID Q9KHF2  
 AC 09KHF2  
 DT 01-OCT-2000 (TREMBREL. 15, Created)  
 DT 01-OCT-2000 (TREMBREL. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBREL. 15, Last annotation update)  
 DE TPRJ  
 OS *Treponema pallidum*.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; *Treponema*.  
 OX NCBI\_TaxID=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SS14;  
 RA Stamm L.V., Bergen H.L.;  
 RA Stamm L.V., Bergen H.L.;  
 RT "Intrastrain Heterogeneity of the TpRK Protein of the *Treponema*  
*pallidum* Nichols Strain and Street Strain 14.;"  
 RT Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF261076; AAP76885.1;  
 SQ SEQUENCE 758 AA; 81285 MW; E4B882007BBF521B CRC64;

Query Match 10.4%; Score 7; DB 2; Length 758;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GRGGGLGG 8  
 |||||  
 Db 169 GRGGGLGG 175

Search completed: July 12, 2001, 06:22:54  
 Job time: 345 sec

RESULT 25  
 Q9KHF1 PRELIMINARY; PRT; 758 AA.  
 ID Q9KHF1  
 AC 09KHF1  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SS14;  
 RA Stamm L.V., Bergen H.L.;  
 RA Stamm L.V., Bergen H.L.;  
 RT "Intrastrain Heterogeneity of the TpRK Protein of the *Treponema*  
*pallidum* Nichols Strain and Street Strain 14.;"  
 RT Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF261076; AAP76884.1;  
 SQ SEQUENCE 758 AA; 81334 MW; DF91F1ED0F6CB3D7 CRC64;

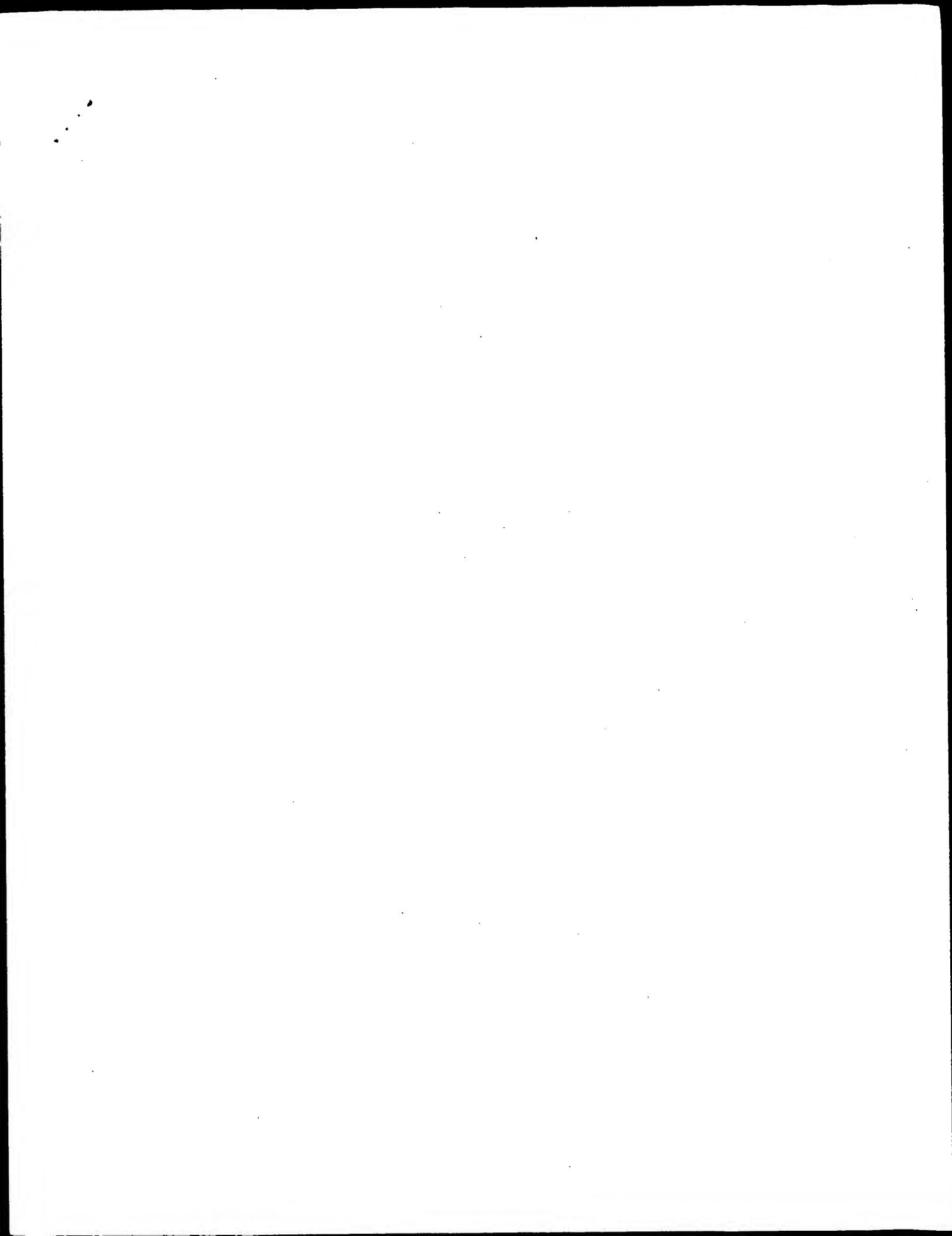
Query Match 10.4%; Score 7; DB 2; Length 758;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGGLGG 8  
 |||||  
 Db 169 GFGGLGG 175

RESULT 25  
 Q9KHF1 PRELIMINARY; PRT; 758 AA.  
 ID Q9KHF1  
 AC 09KHF1

Thu Jul 12 06:30:38 2001

us-09-506-978-1.oligo.rspt



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on:

July 12, 2001, 06:17:34 ; Search time 8.4 Seconds

(without alignments)

273.228 Million cell updates/sec

Title:

US-09-506-978-1

Perfect score:

67

1 GGFGGGGKGKCPSEIFSR.....CRIGYLRNKKKVCPVRSKCG 67

Sequence:

OLIGO

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched:

93435 seqs, 34255486 residues

Word size :

0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : SwissProt\_39;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No. Score Match Length DB ID

Description

RESULT	1
RX	SALA_DROSTI
ID	STANDARD;
RA	P21749; drosophila
AC	"The homeotic gene <i>spalt</i> ( <i>sal</i> ) evolved during <i>Drosophila</i> speciation.";
RL	Proc. Natl. Acad. Sci. U.S.A. 86:5483-5486(1989).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).
DT	01-MAY-1991 (Rel. 18, Last sequence update)
DR	01-FEB-1996 (Rel. 33, Last annotation update)
DE	PROTEIN SPALT-ACCESSORY.
GN	SALA OR SAL
OS	<i>Drosophila simulans</i> (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydioidea; Dirosophilidae; Drosophila.
OX	NCBI_TAXID=7240;
RN	[1]
RP	SEQUENCE FROM N.A.
RP	MEDLINE=89315821; PubMed=2568636;
RA	Reuter, D., Schuh, R., Jaeger, H.;
RT	"The homeotic gene <i>spalt</i> ( <i>sal</i> ) evolved during <i>Drosophila</i> speciation.";
RL	Proc. Natl. Acad. Sci. U.S.A. 86:5483-5486(1989).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).
DT	01-MAY-1991 (Rel. 18, Last sequence update)
DR	01-FEB-1996 (Rel. 33, Last annotation update)
EMBL	M21227; AA28877.1; -
DR	PIR; B33910; B33910;
DR	FlyBase; FBgn0012892; Dsim\sala.
DR	SEQUENCE: 139 AA; 14145 MW; AED5D51561C229CF CRC64;

Query Match	11.9%	Score	8;	DB	1;
Best Local Similarity	100.0%	Pred. No.	0.15;	Length	139;
Matches	8;	Conservative	0;	MSmatches	0;
				Indels	0;
				Gaps	0;
Qy	1 GGFGGG 8				
Db	32 GGFGGG 39				

## ALIGNMENTS

34	6	9.0	738	1	YKF4_YEAST	P35732 saccharomyces_cerevisiae_YKF4
35	6	9.0	770	1	SM4_F_HUMAN	Q95754 homo_sapiens_SM4_F
36	6	9.0	776	1	SM4_F_RAT	Q92143 rattus_norvegicus_SM4_F
37	6	9.0	777	1	PUR2_YARLLI	Q91911 yarrowia_arachidicola_PUR2
38	6	9.0	788	1	TEX_BORPE	Q45388 bordeatella_henriettae_TEX
39	6	9.0	791	1	YY01_MYCLE	Q49736 mycobacterium_smegmatis YY01
40	6	9.0	792	1	PUR2_DROMEE	P07240 dromobacterium_trifunctionale_PUR2
41	6	9.0	802	1	PUR_YEAST	Q07244 s_bifunctional_protein_PUR
42	6	9.0	914	1	WA22_MYCTU	Q06794 mycobacterium_wadsworthii_WA22
43	6	9.0	1005	1	YCF1_OENBE	P31563 oenothera_benjaminii_YCF1
44	6	9.0	1177	1	JAK_DROME	Q24592 drosophila_melanogaster_JAK
45	6	9.0	1353	1	PUR2_DROME	P09667 drosophila_melanogaster_PUR2
46	6	9.0	1364	1	PUR2_DROP5	P16340 drosophila_melanogaster_PUR2
47	6	9.0	1415	1	ICP4_SWMSG	Q03362 marek's_disease_virus_ICP4
48	6	9.0	3075	1	IMAI_HUMAN	P25391 homo_sapiens_IMAI
49	5	7.5	14	1	SAP2_ARBPO	P11760 arbabia_punicea_P22508
50	5	7.5	27	1	FBRL_PHYPO	P02508 physarum_pudicum_FBRL

**HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2591.**  
 RV2591 OR MTC227-10C.  
**OS** Mycobacterium tuberculosis.  
**Bacteria; Firmicutes; Actinobacteria; Actinomycetaceae; Corynebacterineae; Mycobacteriaceae; Mycobacterium.**  
**OC** Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
**NCBI\_TAXID=1773;**  
**[1]** SEQUENCE FROM N.A.  
**RP** STRAIN=R27RV;  
**RX** MEDLINE:98295887; PubMed=9634230;  
**RC** Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D., Gentles S., Hamlin N., Horsnell T.S., Hunt R., Jagerl K., Jones M., Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D., Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V., Walsh S.V., Whitehead S.;  
**RA** Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
**RL** -1- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH SN-RNP U1. IT BINDS  
**CC** STEM LOOP I OF U1 SNRNA.  
**CC** -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
**RA** Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Horstby T., Jagerl K., Krogh A., McLennan J., Moule S., Murphy L., Oliver S., Osborn J., Quail M.A., Rajandream M.A., Rogers J., Rutier S., Seeger K., Skelton S., Squires S., Squires R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
**RT** "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";  
**RL** Nature 393:537-544(1998).  
**CC** -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS  
**CC** SUBFAMILY.  
**CC** This SWISS-PROT entry is copyright. It is produced through a collaboration between the European Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).  
**CC** DR EMBL; Z38005; CRA86162.1; -.  
**CC** DR EMBL; X59966; CAA42602.1; -.  
**CC** DR InterPro; IPR000504; -.  
**CC** DR PIR; S16815; S16815.  
**CC** DR PIR; S48418; S48418.  
**CC** DR B2SP; P1939; 2SXL.  
**CC** DR TRANSFAC; T01242; -.  
**CC** DR SGD; S0001323; SNP.  
**CC** DR InterPro; IPR000504; -.  
**CC** DR Pfam; PF00076; rrm; 1.  
**CC** DR PROSITE; PS50102; RRM; 1.  
**CC** DR PROSITE; PS00300; RRM\_RNP\_1; 1.  
**CC** DR Nucleic acid binding protein; Ribonucleoprotein; RNA-binding.  
**CC** FT DOMAIN 107 198 RNA-BINDING (RNM).  
**CC** SQ SEQUENCE 543 AA; 46287 MW; 59730339E5D2DF59 CRC64;  
**CC** Query Match 11.9%; Score 8; DB 1; Length 543;  
**CC** Best Local Similarity 100.0%; Pred. No. 0.42; Mismatches 0; Indels 0; Gaps 0;  
**CC** Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
**Qy** 1 GGFGGG 8  
**Db** 380 GGFGGGG 387  
**RESULT 3**  
**RV17\_YEAST** STANDARD; PRT; 300 AA.  
**Qy** 1 GGGGGG 8  
**Db** 380 GGFGGGG 387  
**RESULT 4**  
**DHSO\_SCPO** STANDARD; PRT; 360 AA.  
**ID** DHSO\_SCPO  
**AC** P36624;  
**DT** 01-JUN-1994 (Rel. 29 Created)  
**DT** 01-OCT-2000 (Rel. 40 Last annotation update)  
**DT** 01-OCT-2000 (Rel. 40 Last annotation update)  
**DE** PUTATIVE SORBITOL DEHYDROGENASE (EC 1.1.1.14) (L-IDITOL 2-DEHYDROGENASE) (TM1 PROTEIN).  
**GN** TM1 OR SP01773.05C.  
**OS** Schizosaccharomyces pombe (Fission yeast);  
**OC** Eukaryota; Fungi; Ascomycota; Schizosaccharomyces;  
**OC** Schizosaccharomyces pombe (Fission yeast);  
**OC** Schizosaccharomyctes; Schizosaccharomyctaceae;  
**OX** NCBI\_TAXID=4896;  
**RN** [1] SEQUENCE FROM N.A.  
**RP** STRAIN=R27;  
**RC** RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Churcher C.M., Connor R., Copsey T.M., Dear S., Devlin K., Fraser A., Gentles S., Hamlin N., Horsnell T.S., Hunt R., Jagerl K., Jones M., Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D., Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V., Walsh S.V., Whitehead S.;  
**RA** Submitted (NOV-1998) to the EMBL/GenBank/DDJB databases.  
**RP** SEQUENCE OF 5-560 FROM N.A.  
**RC** STRAIN=R27;  
**RC** RA Lyne M., Rajandream M.A., Barrell B.G., Beck A., Borzym K., Klages S., Langer I., Reinhardt R.;  
**RC** RA Submitted (NOV-1998) to the EMBL/GenBank/DDJB databases.  
**RN** [2] SEQUENCE OF 5-560 FROM N.A.  
**RC** STRAIN=R27;  
**RC** MEDLINE:94039112; PubMed=8223615;  
**RC** RA Wagner P., Grimaldi M., Jenkins J.R.;  
**RT** "Putative dehydrogenase tm1 suppresses growth arrest induced by a p53 tumour mutant in fission yeast";  
**RL** Eur. J. Biochem. 217:731-736(1993).  
**RC** SEQUENCE FROM N.A.  
**RP** STRAIN=S288C / AB972;

CC -!\* FUNCTION: SUPPRESSES GROWTH ARREST INDUCED BY A p53 TUMOR MUTANT  
 CC IN FISSION YEAST.  
 CC DR U00016; AAC17167\_1; -.  
 CC HSSP; P25685; 1HDJ.  
 CC DR InterPro; IPR001305; -.  
 CC DR InterPro; IPR001623; -.  
 CC DR InterPro; IPR002939; -.  
 CC DR InterPro; IPR03095; -.  
 CC DR Pfam; PF00226; Dnaj; 1.  
 CC DR Pfam; PF01556; Dnaj\_C; 1.  
 CC DR PRINTS; PRO0625; DNAJPROTEIN.  
 CC DR PROSITE; PS00636; Dnaj\_1; FALSE\_NEG.  
 CC DR PROSITE; PS00637; DNaj\_CXXGXG; FALSE\_NEG.  
 CC DR Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.  
 CC DR DOMAIN 4 68 J\_DOMAIN.  
 CC DR DOMAIN 80 111 GLY-RICH.  
 CC DR REPEAT 141 148 CXXGXKG MOTIF.  
 CC DR REPEAT 158 165 CXXGXKG MOTIF.  
 CC DR REPEAT 184 191 CXXGXKG MOTIF.  
 CC DR REPEAT 198 205 CXXGXKG MOTIF.  
 CC DR METAL 141 141 ZINC 1 (BY SIMILARITY).  
 CC DR METAL 144 144 ZINC 1 (BY SIMILARITY).  
 CC DR METAL 158 158 ZINC 2 (BY SIMILARITY).  
 CC DR METAL 161 161 ZINC 2 (BY SIMILARITY).  
 CC DR METAL 184 184 ZINC 2 (BY SIMILARITY).  
 CC DR METAL 187 187 ZINC 2 (BY SIMILARITY).  
 CC DR METAL 198 198 ZINC 1 (BY SIMILARITY).  
 CC DR METAL 201 201 ZINC 1 (BY SIMILARITY).  
 CC DR SEQUENCE 360 AA; 38851 MW; ADEBB3DDBA163224C CRC64;  
 CC DR SEQUENCE 378 AA; 40399 MW; 723D08BPF62153 CRC64;

Query Match 10.4%; Score 7; DB 1; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 3.5; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; . Mismatches 0; Indels 0; Gaps 0;

Qy	44	PGCVCL 50	93	PGCVCL 99
----	----	-----------	----	-----------

RESULT 5  
 DNJ2\_MYCTE STANDARD; PRT; 378 AA.

AC 049762; DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DE CHAPERONE PROTEIN DNJ2.  
 DE DNJ2 OR B1937\_F2\_56.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Robison K., Smith D.R.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases. STIMULATES, JOINTLY WITH GRPE,  
 CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,  
 CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).  
 CC -!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 10.4%; Score 7; DB 1; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 3.7; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; . Mismatches 0; Indels 0; Gaps 0;

Qy	1	6GFGGGLG 7	82	GGGGGLG 88
----	---	------------	----	------------

RESULT 6  
 DNJ2\_MYCTU STANDARD; PRT; 382 AA.

AC 005825; DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DR 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE CHAPERONE PROTEIN DNJ2.  
 DE DNJ2 OR RV2373C OR MTCY7\_07.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=177;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PUBMED=96634230;  
 RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Elmlander K., Gas S., Barry C.E. III, Tekai F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moulé S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger S., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";  
 RL Nature 393:537-544(1998)  
 CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,  
 CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).  
 CC -!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).

CC !- SIMILARITY: BELONGS TO THE DNAJ FAMILY.  
 CC -! SIMILARITY: CONTAINS 1 J DOMAIN.  
 CC -! SIMILARITY: CONTAINS 1 CIR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: Z95208; CAB08479.1; -. DR HSSP; P25685; 1HDJ; DR Tuberculist; RY2373c; -. DR Interpro; IPR001305; -. DR Interpro; IPR01623; -. DR Interpro; IPR002939; -. DR Interpro; IPR003095; -. DR pfam; PF00226; Dnaj; 1. DR pfam; PF00684; Dnaj\_CXXCNGX; 1. DR PRINTS; PRO0625; DNAPROTEIN. DR PROSITE; PS00636; DNAJ\_1; FALSE\_NEG. DR PROSITE; PS00076; DNAJ\_2; 1. DR PROSITE; PS00637; DNAJ\_CXXCNG; FALSE\_NEG. DR Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding. FT DOMAIN 4 68 FT DOMAIN 7 110 FT REPEAT 145 152 FT REPEAT 162 169 FT REPEAT 188 209 FT METAL 145 148 FT METAL 162 165 FT METAL 188 188 FT METAL 191 191 FT METAL 202 205 SQ SEQUENCE 382 AA; 40489 MW; F04593AB546D57F2 CRC64;

Query Match 10.4%; Score 7; DB 1; Length 382; Best Local Similarity 100.0%; Pred. No. 3.7; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFRGGLG 7 Db 82 GGFGGLG 88

RESULT 7 SP54\_MTCPN ID SP54\_MTCPN STANDARD; PRT; 450 AA.

AC PT5054; CC 01-NOV-1997 (Rel. 35, Last sequence update) DT 01-OCT-2000 (Rel. 40, Last annotation update) DE SIGNAL RECOGNITION PARTICLE PROTEIN (FIFTY-FOUR HOMOLOG). FEH OR MPN061 OR MP093. OS Mycoplasma pneumoniae. OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes; OC Mycoplasmataceae; Mycoplasma. OC NCBI\_TaxID:2104; RN [1] SEQUENCE FROM N.A.

RP STRAIN=ATCC 29342; / M129; RX MEDLINE=97105885; PubMed=8948633; RA Himmelreich R., Hilbert H., Plagens H., Pirkle E., Li B.-C., RA Hermann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae." RT Nucleic Acids Res. 24:4420-4449(1996). EXPRT OF EXTRA-CYTOPLASMIC CC FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE CC RIBOSOMES (BY SIMILARITY).

CC -! SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL CC CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN FFH (BY SIMILARITY). CC BINDS GTP; THE N-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL CC SEQUENCE.

CC !- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS. CC EMBL: AE000011; AAB95741.1; -. DR HSSP; 007347; 1FFH. DR Interpro; IPR003097; -. DR Pfam; PP00448; SRP54; 1. DR PROSITE; PS00300; SRP4; 1. DR Signal recognition particle; GTP-binding; RNA-binding. KW DOMAIN 1 293 FT DOMAIN 294 450 FT NP-BIND 106 113 FT NP-BIND 188 192 FT NP-BIND 246 249 FT SEQUENCE 450 AA; 50141 MW; 7EFE42ECEA4F679 CRC64;

Query Match 10.4%; Score 7; DB 1; Length 450; Best Local Similarity 100.0%; Pred. No. 4.2; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGGLGG 8 Db 441 GFGLGG 447

RESULT 8 K1CL\_HUMAN ID K1CL\_HUMAN STANDARD; PRT; 622 AA.

AC P35527; Q14655; CC 01-JUN-1994 (Rel. 29, Last sequence update) DT 01-JUN-1994 (Rel. 29, Last sequence update) DT 15-JUL-1998 (Rel. 36, Last annotation update) DE KERATIN, TYPE I CYTOSKELETAL 9 (CYTOKERATIN 9) (K9) (CK 9). GN KRT9.

OS Homo sapiens (Human). OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrini; Hominoidea; Homo. OC NCBI\_TaxID:9606;

RN [1] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RP TISSUE=foot sole tissue; RX MEDLINE=94131202; PubMed=7507869; RA Langbein L., Held H.W., Moll I., Franke W.W.; RT "Molecular characterization of the body site-specific human epidermal cyrtokeratin 9: cDNA cloning, amino acid sequence, and tissue specificity of gene expression."; RT Differentiation 55:57-72(1993). RL

RN [2] SEQUENCE OF 449-455. RX MEDLINE=90267446; PubMed=2140676; RA Rosen E.M., Meronky L., Romero R., Setter E., Goldberg I.; RT "Human Placenta contains an epithelial scatter protein."; RL Biochem. Biophys. Res. Commun. 168:1082-1088(1990). RN [3] VARIANTS EPPK LIS-160; GLN-162 AND TRP-162.

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae."

		Query Match	Best Local Similarity	Score	DB	Length
RA	Reis A.; Hennies H.-C.; Langbein L.; Digweed M.; Mischi D.,	10.4%	100.0%	7;	1;	622;
RA	Dressler M.; Schrock E.; Royer-Pokora B.; Franke W.W.; Sperling K.,			No.	5.4;	
RA	Kuester W.;			Mismatches	0;	
RT	"Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK)."; Nat. Genet. 6:174-179(1994).			Indels	0;	Gaps
RL	RN [4]	VARIANTs EPPK TRP-162 AND SER-167.				
RX	RN [4]	VARID=95164983; PubMed=7532199;				
RA	RT Rothnagel J.A.; Wojciech S.; Liefer K.M.; Dominey A.M.; Huber M.,					
RA	RT Hohi D.; Roop D.R.;					
RT	"Mutations in the Ia domain of keratin 9 in patients with epidermolytic palmoplantar keratoderma."; J. Invest. Dermatol. 104:430-433(1995).					
CC	CC -!- FUNCTION: MAY SERVE AN IMPORTANT SPECIAL FUNCTION EITHER IN THE MATURE PALMAR AND PLANTAR SKIN TISSUE OR IN THE MORPHOGENIC PROGRAM OF THE FORMATION OF THESE TISSUE.					
CC	CC -!- SUBUNIT: HETEROCTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.					
CC	CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE TERMINALLY DIFFERENTIATED EPIDERMIS OF PALMS AND SOLES.					
CC	CC -!- DISEASE: DEFECTS IN KRT9 ARE A CAUSE OF EPIDERMOLYTIC PALMOPLANTR KERATODERMA (EPPK), AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY DIFFUSE THICKENING OF THE EPIDERMIS ON THE ENTIRE SURFACE OF PALMS AND SOLES SHARPLY BORDERED WITH ERYTHEMATOUS MARGINS.					
CC	CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].					
CC	CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.					
CC	CC -!- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE A 60 kDa CHAIN OF PLACENTAL SCATTER PROTEIN.					
CC	CC -!- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).					
DR	DR EMBL; X75015; CAA52924; 1; -.					
CC	CC EMBL; S69310; AAC60519; 1; -.					
DR	DR PROSITE; PS00226; IF; 1					
KW	KW Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin; Disease mutation.					
FT	FT DOMAIN 1 151 HEAD.					
FT	FT DOMAIN 152 460 ROD.					
FT	FT DOMAIN 152 622 TAIL.					
FT	FT DOMAIN 152 187 COIL 1A.					
FT	FT DOMAIN 188 205 LINER 1.					
FT	FT DOMAIN 207 298 COIL 1B.					
FT	FT DOMAIN 299 321 LINER 12.					
FT	FT DOMAIN 322 450 COIL 2.					
FT	FT DOMAIN 14 20 POLY-GLY.					
VARIANT	VARIANT 160 160 N -> K (IN EPPK).					
FT	FT /FTid=VAR_003822; R -> Q (IN EPPK).					
VARIANT	VARIANT 162 162 /FTid=VAR_003823; R -> W (IN EPPK).					
FT	FT /FTid=VAR_003824; L -> S (IN EPPK).					
FT	FT /FTid=VAR_003825; T -> SR (IN CAA82215).					
SEQUENCE	SEQUENCE 622 AA; 61987 MN; 898C3B25D4B5ED94 CRC64;					
QV	QV 2 GFGSGG 8					
RP	RP 111111 135					
RESULT	RESULT 9					
ID	ID SCXP_ANMDA					
AC	AC P01498;					
DT	DT 21-JUL-1986 (Rel. 01, Created)					
DT	DT 21-JUL-1996 (Rel. 01, Last sequence update)					
DT	DT 01-NOV-1995 (Rel. 32, Last annotation update)					
DE	DE NEUROTOXIN P2.					
OS	OS Androctonus mauretanicus (Scorpion).					
OC	OC Eukaryota; Metazoa; Arthropoda; Chelicera; Arachnida; Scorpiones; Buthoidea; Buthidae; Androctonus.					
OX	OX NCBI_TAXID=6660;					
RN	RN [1]					
RP	RP SEQUENCE.					
RX	RX MEDLINE=85193276; PubMed=3392595;					
RA	RA DR Rosso J.P.; Rochat H.;					
RT	RT "Characterization of ten proteins from the venom of the Moroccan scorpion Androctonus mauretanicus, six of which are toxic to the mouse."; Toxicon 23:113-125(1985).					
RT	RT PIR; A01758; NTSRM.					
DR	DR HSSP; P15222; ISIS.					
KW	KW Neurotoxin.					
FT	FT DISULFID 1 18 BY SIMILARITY.					
FT	FT DISULFID 4 25 BY SIMILARITY.					
FT	FT DISULFID 15 30 BY SIMILARITY.					
FT	FT DISULFID 19 32 BY SIMILARITY.					
SEQUENCE	SEQUENCE 35 AA; 3673 MW; 213E69262289BB5A CRC64;					
CC	CC -!- SIMILARITY: BELONGS TO THE SHORT SCORPION TOXIN FAMILY.					
DR	DR PIR; A01758; NTSRM.					
DR	DR HSSP; P15222; ISIS.					
QV	QV 7 GGRGKC 12					
Db	Db 111111 25					
RESULT	RESULT 10					
YMC3_OENBE	YMC3_OENBE					
ID	ID YMC3_OENBE					
AC	AC P08747;					
DT	DT 01-AUG-1988 (Rel. 08, Created)					
DT	DT 01-AUG-1988 (Rel. 08, Last sequence update)					
DT	DT 01-FEB-1996 (Rel. 33, Last annotation update)					
DE	DE HYPOTHETICAL 9.2 kDa PROTEIN IN COXIII REGION.					
OS	OS Oenothera berteroana (Bertero's evening primrose).					
OG	OG Mitochondrion.					
OC	OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; core eudicots; Rosidae; eurosids II; Myrtales; Onagraceae; Oenothera.					
OC	OC NCBI_TAXID=3950;					
RN	RN [1]					
RP	RP SEQUENCE FROM N.A.					
RC	RC SPRAIN-CV; MONZIA;					
RA	RA Rothnagel R.; Schopel W.; Schuster W.; Brennicker A.;					
RT	RT "The cytochrome oxidase subunit I and subunit III genes in Oenothera mitochondria are transcribed from identical promoter sequences.";					
RL	RL EMBO J. 6:29-34(1987).					
CC	CC -!- MISCELLANEOUS: POSITION 54 IS MODIFIED BY RNA EDITING.					



CC -!- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-  
 CC ACYL GROUPS IN 3-SN-PHOSPHOGLYCERIDES.  
 CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLYCHOLINE + H(2)O = 1-ACYLGLYCERYL-  
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.  
 CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY. MAY REPRESENT  
 CC A NEW SUBTYPE OF GROUP II PLA2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U07798; AAQ57473.1; ALT\_INIT.  
 DR HSSP; P14418; IPST.  
 DR InterPro; IPR001211; -.  
 DR Pfam; PF00088; phosphatidyl; 1.  
 DR PROSITE; PS00118; PA2\_HIS; 1.  
 DR PROSITE; PS00119; PA2\_ASP; 1.  
 KW Hydrolase; Lipid degradation; Signal; Calcium.  
 FT SIGNAL 1 20  
 FT CHAIN 21 150  
 FT ACT\_SITE 67 67 BY SIMILARITY.  
 FT CA\_BIND 68 68 BY POTENTIAL.  
 FT ACT\_SITE 115 115 BY SIMILARITY.  
 FT DISULFID 46 143 BY SIMILARITY.  
 FT DISULFID 48 64 BY SIMILARITY.  
 FT DISULFID 63 121 BY SIMILARITY.  
 FT DISULFID 69 150 BY SIMILARITY.  
 FT DISULFID 70 114 BY SIMILARITY.  
 FT DISULFID 79 107 BY SIMILARITY.  
 FT DISULFID 97 112 BY SIMILARITY.  
 FT CARBOHYD 92 92 N-LINKED (GLCNAC- .) (POTENTIAL).  
 SQ SEQUENCE 150 AA; 16919 MW; F805/5BET19A3B/FF CRC64;  
 -----  
 Query Match 9.0%; score 6; DB 1; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 GLGGRG 10  
 Db 49 GLGGRG 54  
 -----  
 RESULT 14  
 MAUML\_METEX STANDARD; PRT; 220 AA.  
 AC 049130;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE METHYLAMINE UTILIZATION FERREDOXIN-TYPE PROTEIN MAUML.  
 GN MAUML.  
 OS Methylobacterium extorquens.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Methylomonas group; Methylbacterium.  
 OX NCBI\_TAXID=408;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RN STRAIN=AM1 / NCIB 9133;  
 RN MEDLINE=94229425; PubMed=8021187;  
 RA Chistosarov A.Y., Chistoserdova L.V., McIntire W.S., Lidstrom M.E.;  
 RT "Genetic organization of the mau gene cluster in Methylomonas extorquens AM1: complete nucleotide sequence and generation and characteristics of mau mutants."  
 RT J. Bacteriol. 176:4052-4065(1994).  
 CC -!- FUNCTION: INVOLVED IN ELECTRON TRANSFER (POTENTIAL).  
 CC -!- PATHWAY: METHYLAMINE UTILIZATION  
 CC -----  
 CC SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF

CC -----  
 CC 'BACTERIAL-TYPE' 4Fe-4S FERREDOXINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L26406; AAB43941.1; -.  
 DR HSSP; P46797; JVW.  
 DR InterPro; IPR0011450; -.  
 DR Pfam; PF00037; fer4; 2.  
 DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 1.  
 DR Electron transport; Iron-sulfur; 4Fe-4S.  
 KW METAL 60 60 IRON-SULFUR 1 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 63 63 IRON-SULFUR 1 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 66 66 IRON-SULFUR 1 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 70 70 IRON-SULFUR 1 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 98 98 IRON-SULFUR 2 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 101 101 IRON-SULFUR 2 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 106 106 IRON-SULFUR 2 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 110 110 IRON-SULFUR 2 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 138 138 IRON-SULFUR 3 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 146 146 IRON-SULFUR 3 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 149 149 IRON-SULFUR 3 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 153 153 IRON-SULFUR 3 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 182 182 IRON-SULFUR 4 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 185 185 IRON-SULFUR 4 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 188 188 IRON-SULFUR 4 (4Fe-4S) (BY SIMILARITY).  
 FT METAL 192 192 IRON-SULFUR 4 (4Fe-4S) (BY SIMILARITY).  
 SQ SEQUENCE 220 AA; 23308 MW; EDB3CF81BE2947D9\_CRC64;  
 -----  
 Query Match 9.0%; score 6; DB 1; Length 220;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 GLGGRG 10  
 Db 207 GLGGRG 212  
 -----  
 RESULT 15  
 SURE\_HAEN STANDARD; PRT; 249 AA.  
 ID SURE\_HAEN  
 AC P45681;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE STATIONARY-PHASE SURVIVAL PROTEIN SURE\_HOMOLOG.  
 DE SURE OR H10702.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TAXID=727;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RN STRAIN=RD / K420 / ATCC 51907;  
 RN MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., FitzHugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Heidelberg J., Cotton M.D.,  
 RA Utterback T.R., Haines M.C., Nguyn D.M., Sudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 RT influenzae Rd";  
 RL Science 269:496-512(1995).

CC -I- FUNCTION: NOT KNOWN; COULD BE A PHOSPHATASE.  
 CC -I- SIMILARITY: BELONGS TO THE SURE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL; U32753; AAC22361.1; -.  
 DR TIGR; HIT072; -.  
 DR InterPro; IPR00828; -.  
 DR Pfam; PF01975; Sure; 1.  
 DR PROSITE; PS00595; RIBOSOMAL\_S5; 1.  
 KW Ribosomal protein; Repeat.  
 SQ SEQUENCE 249 AA; 27340 MW; D10F280C95266757 CRC64;

Qy 47 VCRGLY 52  
 Db 172 VCRGLY 177

RESULT\_16

ID	RS2_MOUSE	STANDARD;	PRT;	293 AA.
AC	P25444;			
DR	01-MAY-1992 (Rel. 22, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DE	01-NOV-1997 (Rel. 35, Last annotation update)			
GN	RPS2 OR RPS4 OR LLREP3.			
OS	Mus musculus (Mouse);			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TAXID=10116;			
RN	[1] SEQUENCE FROM N.A., AND SEQUENCE OF 111-128.			
RP	STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;			
RA	MEDLINE-92041821; PubMed=1939063;			
Suzuki K., Olivera J., Wool I.G.;				
RT	"Primary structure of rat ribosomal protein S2. A ribosomal protein with arginine-glycine tandem repeats and RGG motifs that are associated with nucleolar localization and binding to ribonucleic acids."			
RT	J. Biol. Chem. 265:2007-2010(1991)			
RL	-I- SIMILARITY: BELONGS TO THE SSP FAMILY OF RIBOSOMAL PROTEINS.			
CC	CC this SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).			
CC	CC DR EMBL; X57432; CAA40679.1; -.			
DR	PIR; S18828; R31852.			
DR	HSSP; P02357; 1PKB.			
DR	InterPro; IPR00821; -.			
DR	Pfam; PF00333; Ribosomal_S5; 1.			
DR	PROSITE; PS00585; RIBOSOMAL_S5; 1.			
KW	Ribosomal protein; Repeat.			
FT	DOMAIN 9 17 3 X 3 AA TANDEM REPEATS.			
FT	DOMAIN 22 29 2 X 4 AA TANDEM REPEATS.			
FT	DOMAIN 34 52 9 X 2 AA TANDEM REPEATS.			
SQ	SEQUENCE 293 AA; 31231 MW; 9092DB564RA624C9 CRC64;			

Query Match 9.0%; Score 6; DB 1; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GLGGRG 10  
 Db 18 GLGGRG 23

RESULT\_17

ID	RS2_RAT	STANDARD;	PRT;	293 AA.
AC	P27952;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	01-FEB-1995 (Rel. 31, Last annotation update)			
DE	40S RIBOSOMAL PROTEIN S2.			
GN	RPS2.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TAXID=10116;			
RN	[1] SEQUENCE FROM N.A., AND SEQUENCE OF 111-128.			
RP	STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;			
RA	MEDLINE-92041821; PubMed=1939063;			
Suzuki K., Olivera J., Wool I.G.;				
RT	"Primary structure of rat ribosomal protein S2. A ribosomal protein with arginine-glycine tandem repeats and RGG motifs that are associated with nucleolar localization and binding to ribonucleic acids."			
RT	J. Biol. Chem. 265:2007-2010(1991)			
RL	-I- SIMILARITY: BELONGS TO THE SSP FAMILY OF RIBOSOMAL PROTEINS.			
CC	CC this SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).			
CC	CC DR EMBL; X57432; CAA40679.1; -.			
DR	PIR; S18828; R31852.			
DR	HSSP; P02357; 1PKB.			
DR	InterPro; IPR00821; -.			
DR	Pfam; PF00333; Ribosomal_S5; 1.			
DR	PROSITE; PS00585; RIBOSOMAL_S5; 1.			
KW	Ribosomal protein; Repeat.			
FT	DOMAIN 9 17 3 X 3 AA TANDEM REPEATS.			
FT	DOMAIN 22 29 2 X 4 AA TANDEM REPEATS.			
FT	DOMAIN 34 52 9 X 2 AA TANDEM REPEATS.			
SQ	SEQUENCE 293 AA; 31231 MW; 9092DB564RA624C9 CRC64;			

Query Match 9.0%; Score 6; DB 1; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GLGGRG 10  
 Db 18 GLGGRG 23

RESULT\_18

ID	DNAJ_DEIPR	STANDARD;	PRT;	307 AA.
AC	O34136;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			

Query Match 9.0%; Score 6; DB 1; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC

DR EMBL; X83413; CRA58357.1; -.

DR M68963; AAA65574.1; -.

DR PIR; B36769; B36769; 335 AA; 37877 MW; 177BC8ED39DEE014 CRC64;

SEQUENCE

Query Match 9.0%; Score 6; DB 1; Length 335;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 LGYLRN 55  
||| |||  
Db 11 LGYLRN 16

RESULT 21

ID NU2M\_MYXGL STANDARD; PRT; 348 AA.

AC 021078; 15-JUL-1998 (Rel. 36, created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)

CC ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-FORMYLGLYCINAMIDINE =  
-!- CATALYTIC ACTIVITY: ADP + 5'-PHOSPHORIBOSYL-5-AMINOIMIDAZOLE.  
-!- SUBCELLULAR LOCATION: CHLOROPLAST.

DE NADH OR NAD2.

OS Myxine glutinosa (Atlantic hagfish).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;  
OC Myxinidae; Myxinae; Myxine.  
OX NCBI\_TaxID=7769;

RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=97398704; Pubmed=9254918;

RX Delaire C., Barriel V., Tiller S., Javier P., Gachelin G.;  
RT "The main features of the craniate mitochondrial DNA between the NDI  
and the cox genes were established in the common ancestor with the  
lancelet.";  
RL Mol. Biol. Evol. 14:807-813(1997).

CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+)+ + UBIQUINOL.  
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
INNER MEMBRANE.

CC -!- SIMILARITY: BELONGS TO THE COMPLEX 1 SUBUNIT 2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to license@isb-sib.ch).

CC

DR EMBL; Y05277; CAR70717.1; -.

DR InterPro; IPR001750; -.

DR Pfam; PF00361; oxidoreductase\_1.

DR Oxidoreductase; NAD; UbiQuinone; Mitochondrion; Transmembrane.

SQ SEQUENCE 348 AA; 38788 MW; D0123ADD1CF51EA CRC64;

Query Match 9.0%; Score 6; DB 1; Length 348;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFGL 6  
||| |||  
Db 163 GGFGL 168

RESULT 22

ID PUR5\_ARATH STANDARD; PRT; 355 AA.

AC 005728;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE PHOSPHORIBOSYLFORMYLGLYCINAMIDINE CYCLO-LIGASE, CHLOROPLAST PRECURSOR  
(EC 6.3.3.1) (AIRS) (PHOSPHORIBOSYL-AMINOIMIDAZOLE SYNTHETASE) (AIR  
SYNTHASE).

DE (EC 6.3.3.1) (AIRS) (PHOSPHORIBOSYL-AMINOIMIDAZOLE SYNTHETASE) (AIR  
SYNTHASE).

GN PUR5.

OS Vigna unguiculata (Cowpea).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
OC Fabales; Fabaceae; Papilionoideae; Vigna.  
OX NCBI\_TaxID=3917;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN=CV; VITA 3; TISSUE=Root nodules;  
 RA Smith P.M.C.; Mann A.J.; Hall D.J.; Atkins C.A.;  
 RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: ATP + 5'-ORTHOPHOSPHORYL-5-AMINOMIDAZOLE =  
 CC ADP + ORTHOPHOSPHATE + 5'-PHOSPHORYLOSYL-5-AMINOMIDAZOLE.  
 CC [5]  
 CC -!- PATHWAY: FIFTH STEP IN DE NOVO PURINE BIOSYNTHESIS.  
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -!- SIMILARITY: TO OTHER AIRS FROM BACTERIA AND EUKARYOTES.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; U30895; AAC14578.1; -.  
 DR InterPro; IPR000728; -.  
 DR Pfam; PF0086; AIRS; 1.  
 KW Purine biosynthesis; Ligase; Chloroplast; Transit peptide.  
 FT TRANSIT ?  
 FT CHAIN ?  
 SQ SEQUENCE 388 AA; 40427 MW; 15D0B1127C9EE6 CRC64;

Query Match 9.0%; Score 6; DB 1; Length 388;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAFGGL 6  
 |||||  
 Db 93 GAFGGL 98

RESULT 24  
 KICS\_HUMAN STANDARD; PRT; 400 AA.  
 ID KICS\_HUMAN  
 AC P08727;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE KERATIN, TYPE I CYTOSKELETAL 19 (CYROKERATIN 19) (K19) (CK 19).  
 GN KRT19.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Chordata; Craniata; vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrini; Hominidae; Homo.  
 OX NCBI\_TAXID-9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-89225250; PubMed=2469734;  
 RA Stasiak P.C., Purkis P.E., Leigh I.M., Lane E.B.;  
 RT Keratin 19; predicted amino acid sequence and broad tissue distribution suggest it evolved from keratinocyte keratins.;  
 RT J. Invest. Dermatol. 92:707-716(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE-89096504; PubMed=2447559;  
 RA Stasiak P.C., Lane E.B.;  
 RT Sequence of cDNA coding for human keratin 19.;"  
 RL Nucleic Acids Res. 15:10058-10058(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-88124986; PubMed=2448790;  
 RA Eckert R.L.;  
 RT "Sequence of the human 40-kDa keratin reveals an unusual structure with very high sequence identity to the corresponding bovine proc. natl. Acad. sci. U.S.A. 85:1114-1118(1988).  
 RL [4]  
 RP SEQUENCE FROM N.A.

Query Match 9.0%; Score 6; DB 1; Length 400;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FGGLGG 8  
 Db 15 FGGLGG 20

RESULT 25  
 YFJD\_ECOLI STANDARD; PRT;  
 ID YFJD\_ECOLI  
 AC P37908; P76600; P76601; P77009; 428 AA.

DT 01-OCT-1994 (Rel. 30; Created)  
 DT 15-DEC-1998 (Rel. 37; Last sequence update)  
 DR 01-DEC-2000 (Rel. 40; Last annotation update)  
 DE HYPOTHETICAL 45.4 KDa PROTEIN IN FFH-GRPE INTERGENIC REGION.  
 GN EYJD OR B2612/B2613.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TAXID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=927503;  
 RA Blattner F.R., Plunkett G. III., Bloch C.A., Perna N.T., Burland V.,  
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL science 277:1453-1474 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,  
 RA Ikeno T., Isono S., Itoh T., Kanai K.,  
 RA Kasai H., Kashimoto K., Kim S., Kimura S., Kitagawa M.,  
 RA Kitakawa M., Makino K., Matsuda S., Mikl T., Mizobuchi K., Mori H.,  
 RA Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,  
 RA Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C.,  
 RA Yamamoto Y., Yano M.,  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE OF 322-428 FROM N.A.  
 RC STRAIN=B78;  
 RX MEDLINE=88319942; PubMed=3045760;  
 RA Lipinska B., King J., Ang D., Georgopoulos C.;  
 RT "Sequence analysis and transcriptional regulation of the Escherichia  
 coli grpe gene, encoding a heat shock protein.";  
 RL Nucleic Acids Res. 16:7545-7562(1988).  
 RN [4]  
 RP IDENTIFICATION.  
 RA Rudd K.E.;  
 RL Unpublished observations (AUG-1994).  
 CC --! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC --! SIMILARITY: BELONGS TO THE UFP0053 FAMILY. STRONG, TO  
 CC H. INFLUENZAE H0107.  
 CC -! CAUTION: REF. 1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 CC FRAMESHIFT IN POSITION 199 THAT PRODUCES TWO ORFS.  
 CC -! CAUTION: REF. 3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO NUMEROUS  
 CC FRAMESHIFTS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; AEG00347; AAC75661.1; ALT FRAME.  
 DR EMBL; AE000347; AAC75662.1; ALT FRAME.  
 DR EMBL; D98888; BAA16497.1; ALT\_INIT.  
 DR EMBL; X07863; -; NOT\_ANNOTATED\_CDS.  
 DR Ecogene; EGI2442; YfJD.  
 DR InterPro; IPR002550; -.  
 DR Pfam; PF00571; CBS; 2.  
 DR Pfam; PF01595; DUF11; 1.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 4 24 POTENTIAL.  
 FT TRANSMEM 65 85 POTENTIAL.  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT TRANSMEM 130 150 POTENTIAL.

Query Match 9.0%; Score 6; DB 1; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 54 RNKKKV 59  
 Db 297 RNKKKV 302

Search completed: July 12, 2001, 06:23:08  
 Job time: 334 sec

Thu Jul 12 06:30:37 2001

us-09-506-978-1.oligo.rsp



Copyright (C) 1993 - 2000 Compugen Ltd.

Gencore version 4.5

Run on: July 12, 2001, 06:16:45 ; Search time 14.91 seconds  
 (without alignments)

OM protein - protein search, using sw model

Sequence: I GFGGGGRGKPSNEIFSR. .... . CRLGYLRNKKVPRSKC 67

Scoring table: OLIGO  
 Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
 Maximum DB seq length: 200000000

post-processing: Listing first 50 summaries

result No.	Score	Query Length	DB ID	Description
1	8	11.9	B33910	sal homeotic protein - fruit fly ( <i>Drosophila simulans</i> )
2	8	11.9	T08002	glycine rich prote
3	8	11.9	F70726	hypothetical glyc
4	8	11.9	E70917	hypothetical glyc
5	7	10.4	T45643	hypothetical prote
6	7	10.4	T23158	hypothetical prote
7	7	10.4	S16815	SNPL protein - yea
8	7	10.4	T30023	dihydroorotate deh
9	7	10.4	S35981	L-iditol 2-dehydro
10	7	10.4	S72599	hypothetical prote
11	7	10.4	S72587	probable dnaj2 pro
12	7	10.4	B75563	GTP-binding prote
13	7	10.4	S73419	signal recognition
14	7	10.4	F71351	probable preprotei
15	7	10.4	F70760	probable dinkx prot
16	7	10.4	B70213	hypothetical prote
17	7	10.4	I37945	keratin 9, type I,
18	7	10.4	T03504	probable DNA topo
19	7	10.4	B72899	hypothetical prote
20	7	10.4	F71340	probable tpr prote
21	7	10.4	F71301	probable tpr prote
22	7	10.4	T07909	probable alpha-man
23	7	10.4	S8562	hypothetical prote
24	7	10.4	T28734	hypothetical prote
25	7	10.4	H95597	hypothetical prote
26	6	9.0	35	NTSRPM
27	6	9.0	78	neurotoxin P2 - sc
28	6	9.0	9.0	hypothetical prote
29	6	9.0	79	hypothetical prote
30	6	9.0	89	hypothetical prote

Total number of hits satisfying chosen parameters: 219241

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

RESULT	1
Best Local Similarity	11.9%
Matches	8
Conservative	0;
Mismatches	0;
Indels	0;
Gaps	0;

Query Match

QY	1	GFGGGGG	8
Db	32	GFGGGGG	39

### ALIGNMENTS

RESULT	2
Best Local Similarity	10.0%
Matches	8
Conservative	0;
Mismatches	0;
Indels	0;
Gaps	0;

Query Match

QY	1	GFGGGGG	8
Db	32	GFGGGGG	39

Sequence: Onobrychis viciifolia (common sainfoin)

C;Species: Onobrychis viciifolia (common sainfoin)

C;Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 08-Oct-1999

R;Joseph, R.G.

submitted to the EMBL Data Library, October 1997

A;Description: A cDNA coding for a glycine-rich protein from Onobrychis viciifolia

A;Reference number: Z16281

A;Accession: T08002

A;Status: preliminary; translated from GB/EMBL/DDJB

A;Molecule type: mRNA

A;Residues: 1-151 <JDS>

A;Cross-references: EMBL:AF027686; NID:92565428; PIDN:AAB82000.1; PID:92565429

Query Match 11.9%; Score 8; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 0.39; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFGGLGG 8  
Db 76 GGFGGLGG 83

RESULT 3

F70726 hypothetical glycine-rich protein Rv2591 - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: F70726  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Accession: F70726  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-153 <COL>  
A;Cross-references: GB:Z77724; GB:AL123456; NID:g3261620; PIDN:CAB01283.1; PID:gi478221  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv2591  
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 11.9%; Score 8; DB 2; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.1; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFGGLGG 8  
Db 380 GGFGGLGG 387

RESULT 4

E70917 hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 11-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: E70917  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; PMID:98295987  
A;Accession: E70917  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1329 <COL>  
A;Cross-references: GB:Z5844; GB:AL123456; NID:g3250713; PIDN:CAB09271.1; PID:gi2131046  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv1450c  
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 11.9%; Score 8; DB 2; Length 1329;  
Best Local Similarity 100.0%; Pred. No. 2.3; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFGLGG 8  
Db 1318 GGFGLGG 1325

RESULT 5

T45643 hypothetical protein F13I12.120 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000  
C;Accession: T45643  
R;Choisne, N.; Robert, C.; Brodtler, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;  
submitted to the Protein Sequence Database, November 1999  
A;Reference number: Z23010  
A;Accession: T45643  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-100 <CHO>  
A;Cross-references: EMBL:AL133292  
A;Experimental source: cultivar Columbia; BAC clone F13I12  
C;Genetics:  
A;Map Position: 3  
A;Note: F13I12.120  
C;Superfamily: Arabidopsis thaliana hypothetical protein F13I12.120

Query Match 10.4%; Score 7; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 3.2; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFGLGG 7  
Db 89 GGFGGUG 95

RESULT 6

T23158 hypothetical protein K01A6.4 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Accession: T23158  
R;Cottage, A.  
submitted to the EMBL Data Library, January 1996  
A;Reference number: Z19701  
A;Accession: T23158  
A;Status: preliminary; translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 1-284 <WT>  
A;Cross-references: EMBL:268750; PIDN:CAN92963.1; GSPDB:GN00022; CESP:K01A6.4  
A;Experimental source: clone K01A6  
C;Genetics:  
A;Gene: CESP:K01A6.4  
A;Map Position: 4  
A;Introns: 89/3; 104/3; 126/3  
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 10.4%; Score 7; DB 2; Length 204;  
Best Local Similarity 100.0%; Pred. No. 7.5; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFGLGG 7  
Db 259 GGFGGUG 265

RESULT 7

S16015 SNPL protein - yeast (Saccharomyces cerevisiae)  
N;Alternative names: protein YIL061c  
C;Species: Saccharomyces cerevisiae  
C;Accession: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C;Accession: S16015; S48418  
R;Smith, V.; Barrell, B.G.  
EMBO J. 10, 2627-2634, 1991  
A;Title: Cloning of a yeast U1 snRNP 70K protein homologue: functional conservation o

A; Reference number: S16815; MUID:91330888  
 A; Accession: S16815  
 A; Molecule type: DNA  
 A; Residues: 1-360 <SMI>  
 A; Cross-references: EMBL:AL033389; PIDN:CAA21910\_1; GSPDB:GN00067; SPDB:SPBC1773\_05C  
 A; Experimental source: strain AB972  
 R; Smith, V.  
 Submitted to the EMBL Data Library, September 1994  
 A; Reference number: S48407  
 A; Accession: S48418  
 A; Molecule type: DNA  
 A; Residues: 1-300 <SM2>  
 A; Cross-references: GB:Z47047; EMBL:238060; NID:9603997; PID:9763285; GSPDB:GN00009; MIF  
 A; Experimental source: strain AB972  
 C; Genetics:  
 A; Gene: SGD:SNP1; MIPS:YIL061c  
 A; Cross-references: SGD:S001323; MIPS:YIL061c  
 A; Map position: 9L  
 C; Superfamily: Yeast SNP1 protein; ribonucleoprotein repeat homology  
 C; Keywords: nucleus  
 F;108-178/Domain: ribonucleoprotein repeat homology <RMR2>

Query Match 10.4%; Score 7; DB 1; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 7.8; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GGIGGRG 10  
 |||||  
 Db 201 GGIGGRG 207

RESULT 8

T30523 dihydroorotate dehydrogenase (EC 1.3.99.11) - Trypanosoma cruzi  
 C; Species: Trypanosoma cruzi  
 C; Accession: T30523  
 C; Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000  
 J. Mol. Biol. 285, 149-161, 1999  
 A; Title: Novel organization and sequences of five genes encoding all six enzymes for de  
 A; Reference number: Z20841; MUID:9096912  
 A; Accession: T30523  
 A; Status: preliminary; translated from GB/EMBL/DDJB  
 A; Molecule type: DNA  
 A; Residues: 1-314 <GAO>  
 A; Cross-references: EMBL:AB017765; NID:95509902; PIDN:BA474526\_1; PID:g4210458  
 C; Genetics:  
 A; Note: PYRA  
 C; Superfamily: dihydroorotate oxidase  
 C; Keywords: dihydroorotate oxidase

Query Match 10.4%; Score 7; DB 1; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 9.1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 PGCVCR 50  
 |||||  
 Db 93 PGCVCR 99

RESULT 10

S72599 hypothetical protein B1937\_F2\_56 - Mycobacterium leprae  
 C; Species: Mycobacterium leprae  
 C; Accession: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Aug-1999  
 C; Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Aug-1999  
 R; Smith, D.R.; Robison, K.  
 Submitted to the EMBL Data Library, November 1993  
 A; Reference number: S72580  
 A; Accession: S72599  
 A; Status: preliminary  
 A; Molecule type: DNA  
 A; Residues: 1-378 <SMI>  
 A; Cross-references: EMBL:U00016; NID:9466961; PIDN:AAA17167\_1; PID:g466981  
 C; Genetics:  
 A; Start codon: GTG  
 C; Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology <DNJ>  
 C;4-68/Domain: dnaJ amino-terminal homology

Query Match 10.4%; Score 7; DB 2; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 8.1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GFGGGG 8  
 |||||  
 Db 218 GFGGGG 224

RESULT 9

S35981 L-iditol 2-dehydrogenase (EC 1.1.1.14) - fission yeast (Schizosaccharomyces pombe)  
 N; Alternative names: sorbitol dehydrogenase  
 C; Species: Schizosaccharomyces pombe  
 C; Date: 13-Jan-1995 #sequence\_revision 01-Dec-2000 #text\_change 01-Dec-2000  
 C; Accession: T39670; S3845; S35981  
 R; Lye, M.; Ravidrean, M.A.; Barrell, B.G.; Beck, A.; Borzym, K.; Klages, S.; Langer, I.  
 submitted to the EMBL Data Library, November 1998  
 A; Reference number: Z21847  
 A; Accession: T39670

Query Match 10.4%; Score 7; DB 2; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 9.5; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGG 7  
 |||||  
 Db 82 GGGGG 88

RESULT 11

D30587 probable dnaJ protein - Mycobacterium tuberculosis  
 C; Species: Mycobacterium tuberculosis  
 C; Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C; Accession: D70587

R; Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Butter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A; Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A; Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A; Reference number: A70500; MUID:98295987

A; Accession: D70587

A; Status: preliminary; nucleic acid sequence not shown; translation not shown

C; Genetics:

A; Gene: dnaJ2

C; Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F; 4/68/Domain: dnaJ amino-terminal homology <ONJ>

Query Match 10.4%; Score 7; DB 2; Length 382; Best Local Similarity 100.0%; Pred. No. 9; 6; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12

Qy 1 GGREGLG 7

Db 82 GGFGGLG 88

GTP-binding protein Obg - Deinococcus radiodurans (strain R1)

C; Species: Deinococcus radiodurans

C; Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C; Accession: B75563

R; White, O.; Bisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S., M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A; Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A; Reference number: A75250; MUID:20036896

A; Accession: B75563

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-438 <WHI>

A; Cross-references: GB:AE001871; GB:AB000513; NID:96457740; PIDN:AAF09676.1; PID:9645774

A; Experimental source: strain R1

C; Genetics:

A; Gene: D8084

A; Map position: 1

C; Superfamily: GTP-binding protein obg; translation elongation factor Tu homology

Query Match 10.4%; Score 7; DB 2; Length 438; Best Local Similarity 100.0%; Pred. No. 11; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13

Qy 4 GGLGSGRG 10

Db 120 GGLGGRG 126

Best Local Similarity 100.0%; Pred. No. 11; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 10.4%; Score 7; DB 2; Length 450; Best Local Similarity 100.0%; Pred. No. 11; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15

Qy 5 GIGGRGK 11

Db 427 GIGGRGK 433

probable dlnX protein - Mycobacterium tuberculosis (strain H37RV)

C; Species: Mycobacterium tuberculosis

C; Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C; Accession: F70760

R; Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, R.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A; Status: preliminary; nucleic acid sequence not shown; translation not shown

A; Molecule type: DNA

A; Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A; Reference number: A70500; MUID:98295987

A; Accession: F70760

A; Status: preliminary; nucleic acid sequence not shown; translation not shown

A; Molecule type: DNA

A; Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A; Reference number: A70500; MUID:98295987

A; Cross-references: GB:Z74020; GB:AL123456; NID:93261584; PIDN:CAA98325.1; PID:e24855

A; Experimental source: strain H37RV

A; Reference number: S73327; MUID:97105885

A;Gene: dinx

submitted to the EMBL Data Library, December 1993  
 A;Reference number: S41161  
 A;Accession: S41161

Query Match 10.4%; Score 7; DB 2; Length 463;  
 Best Local Similarity 100.0%; Pred. No. 11; Mismatches 0;  
 Matches 7; Conservative 0; Indels 0; Gaps 0;

Qy 4 GGLGRC 10  
 |||||  
 Db 29 GGLGRC 35

RESULT 16

B70213

hypothetical protein BBA50 - Lyme disease spirochete plasmid A/1P54

C;Species: *Borrelia burgdorferi* (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 28-Jul-2000

C;Accession: B0213

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; White, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, B.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.

A;Reference number: A70100; MUID:98065943

A;Accession: B70213

A;Status: Preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 11465 &lt;KLE&gt;

A;Cross-references: GB:HE00790; NID:92690224; PIDN: AAC66282.1; PID:92690281; TIGR:BBA50

A;Experimental source: strain B31

C;Genetics:

A;Genome: plasmid

C;Superfamily: *Borrelia burgdorferi* hypothetical protein BBA50

RESULT 16

B70213

hypothetical protein BBA50 - Lyme disease spirochete plasmid A/1P54

C;Species: *Borrelia burgdorferi* (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 28-Jul-2000

C;Accession: B0213

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; White, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, B.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.

A;Reference number: A70100; MUID:98065943

A;Accession: B70213

A;Status: Preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 11465 &lt;KLE&gt;

A;Cross-references: GB:HE00790; NID:92690224; PIDN: AAC66282.1; PID:92690281; TIGR:BBA50

A;Experimental source: strain B31

C;Genetics:

A;Genome: plasmid

C;Superfamily: *Borrelia burgdorferi* hypothetical protein BBA50

RESULT 17

B70213

hypothetical protein BBA50 - Lyme disease spirochete plasmid A/1P54

C;Species: *Borrelia burgdorferi* (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 28-Jul-2000

C;Accession: B0213

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; White, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, B.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.

A;Reference number: A70100; MUID:98065943

A;Accession: B70213

A;Status: Preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 11465 &lt;KLE&gt;

A;Cross-references: GB:HE00790; NID:92690224; PIDN: AAC66282.1; PID:92690281; TIGR:BBA50

A;Experimental source: strain B31

C;Genetics:

A;Genome: plasmid

C;Superfamily: *Borrelia burgdorferi* hypothetical protein BBA50

RESULT 17

B70213

hypothetical protein BBA50 - Lyme disease spirochete plasmid A/1P54

C;Species: *Borrelia burgdorferi* (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 28-Jul-2000

C;Accession: B0213

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; White, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, B.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.

A;Reference number: A70100; MUID:98065943

A;Accession: B70213

A;Status: Preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 11465 &lt;KLE&gt;

A;Cross-references: GB:HE00790; NID:92690224; PIDN: AAC66282.1; PID:92690281; TIGR:BBA50

A;Experimental source: strain B31

C;Genetics:

A;Genome: plasmid

C;Superfamily: *Borrelia burgdorferi* hypothetical protein BBA50

RESULT 18

T03504

hypothetical protein BBA50 - Lyme disease spirochete plasmid A/1P54

C;Species: *Rhodobacter capsulatus*

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999

C;Accession: T03504

R;Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Ronstein, M.

Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997

A;Title: Sequence of a 189-kb segment of the chromosome of *Rhodobacter capsulatus* SB1

A;Reference number: 214955; MUID:97404404

A;Status: Preliminary; translated from GB/EMBL/DDJB

A;Molecule type: DNA

A;Residues: 1-652 &lt;VIC&gt;

A;Cross-references: EMBL:AF010496; NID:93128256; PIDN: AAC16157.1; PID:93128305

C;Map position: 1

C;Keywords: DNA topoisomerase (ATP-hydrolyzing) chain B

C;Keywords: isomerase

RESULT 18

T03504

hypothetical protein BBA50 - Lyme disease spirochete plasmid A/1P54

C;Species: *Rhodobacter capsulatus*

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999

C;Accession: T03504

R;Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Ronstein, M.

Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997

A;Title: Sequence of a 189-kb segment of the chromosome of *Rhodobacter capsulatus* SB1

A;Reference number: 214955; MUID:97404404

A;Status: Preliminary; translated from GB/EMBL/DDJB

A;Molecule type: DNA

A;Residues: 1-652 &lt;VIC&gt;

A;Cross-references: EMBL:AF010496; NID:93128256; PIDN: AAC16157.1; PID:93128305

C;Map position: 1

C;Keywords: DNA topoisomerase (ATP-hydrolyzing) chain B

C;Keywords: isomerase

RESULT 19

T28899

hypothetical protein T17H7.1 - *Caenorhabditis elegans*C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C;Accession: T28899

R;Favello, A.

RESULT 19

T28899

hypothetical protein T17H7.1 - *Caenorhabditis elegans*C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C;Accession: T28899

R;Langbein, L.

submitted to the EMBL Data Library, January 1996  
 A; Description: The sequence of *C. elegans* cosmid T17H7.

C; Genetics:  
 A; Gene: TP0621

Query Match 10.4%; Score 7; DB 2; Length 758;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0;  
 A; Status: preliminary; translated from GB/EMBL/DDBJ  
 A; Residues: 1-1682 <FAV>  
 A; Cross-references: EML:U42841; PIDN: AAC48170\_1; GSPDB: GN00021; CESP:T17H7.1  
 A; Experimental source: strain Bristol N2; clone T17H7  
 C; Genetics:  
 A; Gene: CBSP:T17H7.1  
 A; Map position: 3  
 A; Introns: 85/3; 147/3

Query Match 10.4%; Score 7; DB 2; Length 682;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0;  
 Qy 4 GGGGGRG 10  
 Db 587 GGUGRG 593

RESULT 22  
 F71340 Query Match 10.4%; Score 7; DB 2; Length 758;  
 C; Species: Treponema pallidum subsp. pallidum (*syphilis* spirochete)  
 C; Accession: C71340  
 C; Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
 R; Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M  
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A; Title: Complete genome sequence of *Treponema pallidum*, the *syphilis* spirochete.  
 A; Reference number: A71250; MUID: 98332770  
 A; Accession: C71340  
 A; Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A; Molecule type: DNA  
 A; Residues: 1-1756 <COL>  
 A; Cross-references: GB:AE001211; GB:AE000520; NID:93322582; PIDN: AAC65300\_1; PID:9332  
 A; Experimental source: strain Nichols  
 C; Genetics:  
 A; Gene: TP0317

Query Match 10.4%; Score 7; DB 2; Length 682;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0;  
 Qy 2 GFGGUGG 8  
 Db 169 GFGGUGG 175

RESULT 23  
 G71133 Query Match 10.4%; Score 7; DB 2; Length 762;  
 C; probable alpha-mannosidase - *Pyrococcus horikoshii*  
 C; Species: Pyrococcus horikoshii  
 C; Accession: G71133  
 R; Kawarabayasi, Y.; Swada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se  
 M.; Ohfuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Ogu  
 DNA Res. 5, 55-76, 1998  
 A; Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
 A; Reference number: A71000; MUID: 9844137  
 A; Accession: G71133  
 A; Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A; Molecule type: DNA  
 A; Residues: 1-856 <KAN>  
 A; Cross-references: GB:AP00003; NID:93236130; PIDN: BAA29929\_1; PID:d1030872; PID:932  
 A; Note: this accession replaces an interim accession for a sequence replaced by Genba  
 C; Genetics:  
 A; Gene: PH0835

Query Match 10.4%; Score 7; DB 2; Length 856;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0;  
 Qy 53 LRNNKKV 59  
 Db 713 LRNNKKV 719

Query Match 10.4%; Score 7; DB 2; Length 758;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0;  
 Qy 2 GFGGUGG 8  
 Db 169 GFGGUGG 175

RESULT 21  
 F71301 Query Match 10.4%; Score 7; DB 2; Length 758;  
 C; probable tpr protein J (tprrJ) - *syphilis* spirochete  
 C; Species: *Treponema pallidum* subsp. *pallidum* (*syphilis* spirochete)  
 C; Accession: F71301  
 C; Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
 R; Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M  
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A; Title: Complete genome sequence of *Treponema pallidum*, the *syphilis* spirochete.  
 A; Reference number: A71250; MUID: 98332770  
 A; Accession: F71301  
 A; Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A; Molecule type: DNA  
 A; Residues: 1-1758 <COL>  
 A; Cross-references: GB:AE001237; GB:AE000520; NID:93322916; PIDN: AAC65595\_1; PID:9332291  
 A; Experimental source: strain Nichols

RESULT 24  
T28734 hypothetical protein F26G5.9 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
R;Sammans, L.; Wohldmann, P.; Beck, C.  
submitted to the EMBL Data Library, September 1997  
A;Description: The sequence of C. elegans cosmid F26G5.  
A;Reference number: Z20516  
A;Accession: T28734  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-944 <SAM>  
A;Cross-references: EMBL:AF022974; PIDN:AAC48041.1; GSPDB:GN00023; CESP:F26G5.9  
A;Experimental source: strain Bristol N2; clone F26G5  
C;Genetics:  
A;Gene: CRSP:F26G5.9  
A;Map position: 5  
A;Introns: 12/3; 48/3; 258/3; 327/3; 643/1; 916/2

---

Query Match 10.4%; Score 7; DB 2; Length 944;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 GFGGGGG 8  
Db 76 GFGGGGG 82

RESULT 25  
H96597 hypothetical protein T5A14.5 [imported] - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: H96597  
R;Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Klaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzalali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Schwartz, J.R.; Shinm, P.; Southwick, A.M.; Sun, H.; Tallon, A.; Authors: Salzberg, S.L.; Schwartz, J.R.; Shinm, P.; Southwick, A.M.; Sun, H.; Tallon, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A;Reference number: AB6141; MUID:21016719  
A;Accession: H96597  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1804 <STO>  
A;Cross-references: GB:AE005173; NID:94204261; PIDN:ABD10642.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: T5A14.5  
A;Map position: 1

---

Query Match 10.4%; Score 7; DB 2; Length 1804;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFGGG 7  
Db 1733 GFGGGGG 1739

Search completed: July 12, 2001, 06:17:29  
Job time: 44 sec

